Queensland University of Technology

Essays on Intentions and Social Preferences

Yola Engler

Submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy (Economics)

School of Economics and Finance and QuBE

Queensland University of Technology

Principal Supervisor: Prof. Dr. Uwe Dulleck

Associate Supervisor: Prof. Lionel Page, Ph.D.

2016

Acknowledgements

I would like to thank my supervisor Uwe Dulleck who made the opportunity to pursue the Ph.D. in Brisbane possible. His door was always open, his advice always very valuable. I also want to thank my other supervisor Lionel Page. His energy and excitement for research was catching and motivating. His support was truly exeptional and his excellent comments and ideas inspiring.

I am also extremely grateful to Rudi Kerschbamer. I profited incredibly from his insightful ideas, comments and helpful support. His short-course on social preferences at the very early stages of my Ph.D. set the course for my research of the past three years. His (theoretical) precision and love for details were a great incentivization and made my research more sophisticated.

Moreover, I owe a special thanks to Daniel Mueller and his love for ice hockey. If it were not for him and our talk at the game of the Mannheimer Adler in January 2012, I would most certainly not have come to Australia. I also want to thank him together with my other colleagues and friends at QUT – especially Ambroise, Ann-Kathrin, Ben, Dave, Jonas, Juliana, Marco, Naomi, Osei, Romain, Steve, Suzanne and Tony – for making my time in Australia such a wonderful experience. Particularly worthy of mentioning are additionally Markus whose programming assistance and

technical expertise made my life easier, and last but not least the good soul of the faculty Poli.

I am also grateful for all the valuable discussions with and opinions from many other friends, colleagues, and participants of conferences and seminars. In particular, I want to thank Colin Camerer, Jim Cox, Ben Greiner, Glenn Harrison, Daniel Houser, Martin Kocher, Marie-Claire Villeval and Daniel Zizzo.

Another thank you goes to the members of the Economics departments at the University of Innsbruck and the London School of Economics (LSE), in particular Rudi Kerschbamer and Balazs Szentes, for hosting me as a visiting Ph.D. student. My stay in both faculties was an inspiring time.

Furthermore, I acknowledge the generous financial support from QUT providing me with a research scholarship and travel funds. The latter not only allowed me to write parts of my thesis at the University of Innsbruck and the LSE but also to attend several conferences in Europe and Australia as well as a summer school in Seoul. These experiences including the interactions with colleagues were not only valuable for my research but also for me personally.

Last but not least, I thank my family and especially my parents. Their love and confidence in me was and is invaluable. Their questions and our resulting discussions often challenged me and my work and thereby improved my research. I also want to thank them for their countless emails with newspaper articles and links to TV discussions or documentaries which were constant thought impulses.

Statement of Original Authorship

The work contained in this thesis has not been previously submitted to meet requirements for an award at this or any other higher education institution. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made.

Brisbane, 22. June 2016

Abstract

This thesis presents three essays on social preferences and the influence of the intentions of another party. Each applies conceptual and methodological tools from behavioural and experimental economics. All studies investigate reciprocity and its relevance but do so in different strategical settings. Reciprocity refers to people's predisposition to reward someone perceived as kind and to punish unkind individuals.

The first study examines the role of reciprocity during a bargaining process: Communicating one's preference over the final outcome of the bargaining does impact the willingness of the other agent to strike a deal. It turns out that adopting too tough a bargaining stance can lead to worse outcomes. The second and third study seek to identify the psychological factors driving reciprocal behaviour. Specifically, the second essay investigates complex "revealed intentions" as specific combinations of possible gains and losses for the involved players. The main finding is that intention-based benevolence is more than just repaying another person's generosity. Individuals also react prosocially to the other's willingness to be vulnerable, i.e. his willingness to take the risk of being worse off by acting than by maintaining the status quo. The third essay emphasizes the role of beliefs on an agent's evaluation of another's intention. The hypothesis is that individuals react differently because of their beliefs about other's payoff expectations, but do so systematically in accordance with the theories of guilt aversion and kindness-reciprocity. Results do not support our claim but strengthen the previously often found "no-effect" of expectations on pro-social behaviour.

Keywords

Bargaining; Beliefs; Behavioural economics; Emotions; Experiments; Guilt aversion; Intentions; Negotiations; Other-regarding preferences; Reciprocity; Social norms; Social preferences; Trust; Trustworthiness

Contents

List of Figures

List of Tables

Part I

General Introduction

Chapter 1

Introduction

"There is no duty more indispensable than that of returning a kindness."

Cicero in his first book of "De Officiis" written 44 BC.[1]

"A man ought to be a friend to his friend and repay gift with gift. People should meet smiles with smiles and lies with treachery."

The Edda, a 13$^{\text{th}}$ century collection of Norse epic verse.[2]

Why do people tip in a restaurant? Why do people participate in collective movements such as political demonstrations? And why do they

[1]The original reads "Sin erunt merita, ut non ineunda, sed referenda sit gratia, maior quaedam cura adhibenda est; nullum enim officium referenda gratia magis necessarium est.". The quote refers to the last part and is a translation by C. R. Edmonds (Edmonds, 1855).

[2]The quote is from the translation by D. E. Martin Clarke in *The Hávamál, with Selections from other Poems in the Edda, Illustrating the Wisdom of the North in Heathen Times* (Clarke, 1923).

spend time and effort to write or review Wikipedia articles? More generally: Why do people forgo their own money and time to help or cooperate with strangers even if they will never meet (again)?

Observations of reciprocal behaviour following the motto "tit for tat" as well as altruism towards strangers are not rare in everyday life. This dissertation examines the impact of several intentions and driving forces on an agent's pro-social behaviour. It contains three studies that aim to contribute to the better understanding of reciprocal interactions by applying conceptual and methodological insights from behavioural and experimental economics.

This thesis takes as its point of origin the cognisance that people often care about the well-being and actions of others, and cooperate daily with each other. In fact, many influential economists like Adam Smith, Kenneth Arrow or Amartya Sen made this point and indicated that such behaviour may also have important economic consequences (Fehr and Schmidt, 2006). Nevertheless, economists ignored this intuition for a long time and routinely modelled (and many still model) decision-makers as purely self-interested money-maximizers. That is, economists have assumed that (i) individuals are rational decision-makers who seek to maximize their utility, where (ii) utility is defined in terms of the individual benefit such as own monetary profit. While the material self-interest assumption started as a convenient proxy for other potential motives (Mullainathan and Thaler, 2000), it quickly became a or even *the* key characteristic of the "homo oeconomicus". While Simon (1955) had already challenged the capacity of perfect information processing, and Tversky and Kahneman (1974) put pressure on the rationality assumption, it was not until the 1980s that the selfishness really came under attack. The experiments conducted by Güth et al. (1982) provided some

of the earliest evidence contradicting this standard view: A substantial number of people seem to be strongly motivated by concerns for altruism, fairness, and reciprocity (Fehr and Schmidt, 2006).[3] Slowly, initial scepticism waned as the findings proved to be robust and systematic (Henrich et al., 2005). By now, the evidence gathered on behavioural patterns that cannot be explained by exclusively selfish preferences is overwhelming. Human behaviour is simply more complex. Fairness considerations proved to motivate people's actions especially in experiments in which decisions about payoff allocations among participants are made (Fehr and Schmidt, 2006; Cox et al., 2007) but are also important in bilateral negotiations. Theoretical and empirical work has shown that they also influence outcomes in market settings (e.g. Fehr et al., 1993, 1997).

Economists generally might be tempted to call people's social behaviour irrational. However, Andreoni and Miller (2002) as well as Fisman et al. (2007) show that it can be expressed in the traditional economic language of a well-behaved preference ordering and is thus *rational*. Such "social" or "other-regarding preferences" capture people's valuation for their own material resources but also include a concern, positive or negative, for the (material) well-being of others.[4] Thereby, the rationality assumption is retained but the conception of agents' utility function is modified, thus adjusting the "selfishness" assumption.

But not only economists have struggled to account for the empirical evidence with their models. Cooperation and pro-social behaviour are

[3]Note that there existed evidence on cooperation already beforehand (see Sally, 1995, for a review). Furthermore, a rational choice model of pure altruism by Francis Edgeworth dates back as far as the 19^{th} century as discussed in some of David Collard's work (e.g. Collard, 1975).

[4]In the remainder of the thesis, I will be using the terms social and other-regarding synonymously.

also hard to explain from an evolutionary point of view as they seem to contradict Darwin's principle of "the survival of the fittest" (Darwin, 1969): The theory of natural selection has often been understood as implying that individuals should selfishly promote their own interest. Thus, it may not be surprising that the Science magazine listed the "Evolution of Cooperation" as one of the most fundamental and broad-ranging 25 unsolved puzzles which present an opportunity to be exploited (Pennisi, 2005).

Research on social behaviour has by now spread its wings across many disciplines – starting from psychology to evolutionary biology to economics and even into neuroscience. This provides a unique opportunity for interdisciplinary thought and method exchange and indeed such collaborations led to much progress over the last couple of years as can be seen from the insights already gained from behavioural and neuroeconomics.

1.1 Evolution of pro-social behaviour

Evolutionary biologists define altruistic behaviour as an action that benefits another organism at a cost (Trivers, 1971). In evolution, the essential measurement is "reproductive fitness" or the expected number of offspring. By acting altruistically, the incurring costs reduce the individual's own number of offspring while boosting the other's number (Okasha, 2013).

Altruism is very common throughout the animal kingdom and not peculiar to humans. Monkeys give alarm calls to warn group members of predators although this increases their own risk of being attacked because they attract attention by doing so. Vampire bats are another example:

They often donate collected blood to fellow bats which were unsuccessful in their own nightly hunt to ensure that they do not starve. Nevertheless, the scope and variety of altruism and cooperation observed among humans is quite extraordinary and distinct (Fehr and Fischbacher, 2003; Bowles and Gintis, 2009).

Already Darwin himself acknowledged such pro-social behaviour as a challenge to his theory of natural selection, which teaches us that organisms should behave in ways that boost their own rather than the other's chances of survival and reproduction. His proposed explanation was group selection. Darwin argued in *The Descendent of Man* (1871, p. 166) that "a tribe including many members who [...] were always ready to give help to each other and sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection".[5] However, the major weakness if such a selection process is to work is what Dawkins (1976) calls "subversion from within": Within any group, altruists would be exploited by selfish free-riders who benefit from the altruists without incurring the costs of behaving altruistically themselves. Thus, they would have a fitness advantage.

Hamilton (1964) proposed the alternative kin selection theory. His theory assumes discriminating altruists who only behave altruistically towards relatives who share their own genes. From a gene's perspective, it makes sense to cause its host to act altruistically towards kin who are also (likely) bearers of this gene. In this way, a gene can assure the maximization of its numbers as long as the so-called "Hamilton rule" is satisfied: the benefit received by the recipient has to be bigger than the costs for the donor divided by the coefficient of relationship between the

[5]For an interesting discussion on the importance of ancestral sociality and group-membership on pro-social behaviour, see also Caporael et al. (1989), for instance.

two involved parties which is defined by the probability that they share the same genes. It predicts that the degree of altruism will be greater, the closer the relationship. While kin selection is broadly accepted to play a major role, it still cannot explain altruistic behaviour towards non-relatives which is widely observed.

For unrelated individuals, Trivers (1971) proposed the theory of reciprocal altruism: Individuals help others to the degree that they can anticipate help in return. Pro-social behaviour occurs when the role of donor and recipient of an altruistic act alternates over time. Any costs incurred in a particular interaction are compensated by the benefits received after several repeated transactions. Trivers explained the evolution of cooperation as instances of mutually altruistic acts. Giving is then self-interested as the costs associated with the initial act of giving are outbalanced by future repayments.

A remaining puzzle are the costly traits like "pure altruism" or what is typically called "strong reciprocity" (Bowles and Gintis, 2000). Pure altruism describes the unconditional willingness to make personally costly contributions to others (including strangers). Similarly, strong reciprocity imposes a cost on the giver without the prospect of repayment. However, giving is not unconditional anymore in the sense that it depends on the process resulting in the decision situation: People repay gifts and punish the violation of fairness and/or cooperation norms even in anonymous one-shot encounters where reputation gains are absent. Fehr and Henrich (2003) argue that strong reciprocity is not necessarily just a maladaptation to the previously discussed theories nor the product of the social evolution[6]. However, convincing alternative explanations are still to be

[6]Social evolution refers to our second system of inheritance through cultural transmissions made from generation to generation.

found.

While the ultimate cause(s) of pro-social tendencies in the human nature is still not entirely understood, the widespread occurrence of co-operation, adherence to social norms and altruism most certainly cannot be entirely explained by genes. Instead, us humans developed methods to externalize knowledge and to teach this knowledge to our offspring (El Mouden et al., 2012; Binmore, 2005, p. 13). An increasing population density in hunter-gatherer societies came with an ascending rate of interaction and the rise of a cumulative culture. Evolution, nowadays, occurs less and less on a biological level but rather by cultural adaptation: The invention of technologies or the development of social institutions are examples of how humans are able to circumvent biology and to establish certain behavioural traits bringing pro-sociality to the next level.[7]

1.2 Aim and outline of the thesis

Previous research has shown that most altruistic behaviour is discriminatory and that people do not seek to uniformly help others (Rabin, 1993, 2002). Instead, pro-social behaviour often depends on the previous action(s) of other agents. This dissertation aims to further investigate the role of intentions and their influence on people's pro-sociality towards the other person. Here, strong reciprocity plays a major role: People have the predisposition to reward nice behaviour with kindness and to punish unkindness – even if there is no prospect of gains in the future or reputation effects. "Tit for tat", "What goes around, comes around" or "A Roland for an Oliver" – there exist many expressions in the English

[7] Zizzo (2003), for instance, discusses interdependend preferences where genes and environment codetermine preferences.

language that describe this behavioural principle which is followed by most of us on a daily basis. Everyday-life examples are numerous: the waitress' smile that assures her a larger tip, the dinner invitation that follows fulfilling a favour or consumers who do not buy products from an as unfair perceived firm although the products' material value may be bigger than their price. Furthermore, salespersons but also charitable organizations know about people's internalized norm of reciprocity. A free food sample, for instance, not only exposes customers to the product but also makes them feel indebted, making them more likely to buy the product (Cialdini, 2001).

The experimental evidence confirms the importance of intentions of one person on the succeeding choice of another person: People are frequently willing to bear a personal cost in order to harm people that are perceived as unkind/unfair or to help kind/fair people – even in one-shot situations (Rabin, 1993). This means that reciprocity is not driven by present material benefits nor the expectation of any in the future. The absence of nepotistic motives distinguishes it from cooperative behaviour in repeated interactions (Gintis, 2000; Fehr and Fischbacher, 2003), which can be sustained as an equilibrium even among solely self-interested individuals.[8] Instead strong reciprocity constitutes a powerful source of the extraction of all efficiencies through its incentive to cooperate even in non-repeated encounters with strangers.

Part II of this thesis presents three empirical essays on other-regarding preferences triggered by intentions. While each of these chapters is self-contained, they are all grounded on the methodological and conceptual

[8]The economic term "cooperative behaviour" refers to acts that Trivers (1971) classified under the term "reciprocal altruism".

insights from experimental and behavioural economics. Furthermore, they are interrelated by their contribution to the understanding of the underlying motivational mechanisms driving pro-social behaviour. All three chapters pay special attention to the concept of strong reciprocity[9] but do so in different strategical settings. This thesis aims to provide new insights into the dynamics of and the proximate motivations that are present in sequential interactions. By investigating the concepts of reciprocity, vulnerability-responsiveness, more complex intention-based benevolence (such as trustworthiness) as well as guilt aversion, the presence of prevailing social norms and the role of pride, the research reveals first factors that prevent and that encourage pro-social behaviour. The recognition of these drivers provides the basis for strengthening and enhancing efficiencies in exchanges and negotiations.

The first essay *Driving a hard bargain is a balancing act: The importance of reciprocity in bargaining*[10] investigates the role of reciprocity during a bargaining process. Results from previous alternating-bargaining games, in which two players have to find an agreement on how to divide a given surplus via a process of offers and counter-offers, have indicated that participants do not play as if they only aim to maximize their own payoff. Instead, claims and final outcomes are biased towards an equal split. The main contribution of this study is the recognition of the bargaining *process* as an important determinant of fairness judgements, which in turn are crucial in determining the success or failure of the negotiation. Employing a double-alternating bargaining game without shrinkage of the pie and using a within-subject design, we find strong

[9] In the following, I will neglect the "strong" and only talk about reciprocity. It will be clear from the context that it refers to "strong reciprocity" rather than "reciprocal altruism" as this thesis focuses on one-shot situations. If potential misunderstanding could occur, I will use the exact terms.

[10] This work is co-authored by Lionel Page.

support for intention-based preferences as an important determinant of a bargainer's willingness to strike a (fair) deal. The optimal first proposal from a first mover consists of a request slightly above the equal payoff split. Higher requests will be punished through lower final offers that willingly hazard the consequence of a potential failure of agreement. Our results thereby indicate that the punishment is not solely driven by distributional preferences but that reciprocal concerns create additional boundaries on how tough one should be in order to reach the best outcome in a bargaining process. In that sense, driving a hard bargain is a balancing act.

Chapter 5 presents the research project *Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games*[11]. The novelty of this research project is an approach which allows the investigation of complex "revealed intentions" as specific combinations of possible gains and losses for the involved players. Following the revealed altruism approach developed by Cox et al. (2008a), we study how the observable properties of a first mover's choice influence a second mover's decision to be more or less benevolent towards the first mover. We extend their approach by looking at the effect of gains and *losses* for both players created by the choice of a first mover on the action of a second mover. Our main finding is that intention-based benevolence is not equal to positive reciprocity but can also be influenced through other factors than generosity: We find that besides the possibility of gains for the second mover (generosity), the risk of losses for the first mover (vulnerability) is an important drivers for second mover behaviour. The availability of a deal and an aversion against violating trust, contrariwise, seem to be far less important motivations.

[11]The paper is joint work with Lionel Page and Rudolf Kerschbamer.

Chapter 6 presents the essay *Guilt-averse or reciprocal: Looking at behavioural motivations in the trust game*[12]. It addresses the unsolved question whether the positive back-transfers observed in so called "trust games" (Berg et al., 1995) are driven by (i) unconditional/pure altruism, (ii) guilt aversion (Dufwenberg and Kirchsteiger, 2004), which is the tendency to fulfil others' manifest expectations in order to avoid the feeling of guilt arising from consciously letting others down, or by (iii) reciprocity in the sense of a willingness to repay a kind action even at some cost (e.g. Rabin, 1993; Falk et al., 2005). The aim of this study is to abstain from the idea that there is only *one* correct theory but to test for individual heterogeneities. For this purpose, we use a modified trust game that allows us to disentangle individual behavioural patterns that are motivated by the three channels described above. The beauty of the trust game is that the variation of the trustee's belief about the trustor's payoff expectations results in opposing predictions for guilt-averse and for reciprocal agents (in terms of their pro-social behaviour). If the trustee is motivated by reciprocity, ceteris paribus higher expectations of the trustor should reduce the perceived kindness of his transfer, and thus the trustee will return less. If the trustee, however, is motivated by guilt aversion, the higher expectations of the trustor should induce higher back-transfers. If the trustee is not motivated by either one, but is purely altruistic, his back-transfer will be positive but independent of the trustor's expectations. By varying the probability that the trustee in a trust game gets the chance to reciprocate, we can exogenously manipulate the trustor's expected reward and thus the second-order beliefs of the trustee (Strassmair, 2009). Using a within-subject design allows us to determine individual differences and to disentangle trustees motivated

[12]It is a joint study with Lionel Page and Rudolf Kerschbamer.

by either of the proposed explanations. However, in contrast to our research hypothesis, we find very limited evidence of type heterogeneity. This suggests that the two existing theories may imperfectly explain the second mover behaviour in the trust game.

In the remainder of this introductory part, I will discuss the methodology of laboratory experiments in Chapter 2. Additionally, in Chapter 3, I will give a short overview of the empirical findings and theoretical developments regarding social preferences that are most relevant for my research without imposing the claim to be exhaustive. The literature overview will be complemented by the more detailed and confined literature reviews contained in Chapters 4 to 6 of Part II.

Part II is the "heart" of this thesis and is devoted to my research projects which are presented in three self-contained essays – one per chapter. Chapter 4 presents the research project *Driving a hard bargain is a balancing act: The importance of reciprocity in bargaining.* Chapter 5 is dedicated to the study *Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games*, and the last essay *Guilt-averse or reciprocal: Looking at behavioural motivations in the trust game* is presented in Chapter 6.

Part III concludes with a summary (Chapter 7), a discussion (Chapter 8) and some concluding remarks (Chapter 9).

Chapter 2

Methodology

I chose to study social preferences by developing and conducting laboratory experiments. While they were always a widely used methodology for advancing causal knowledge in physical and life sciences, their adaptation has been much slower in social sciences. In economics, the first lab experiment was only conducted in the late 1940s (Falk and Heckman, 2009). An initial low level of lab experiments however has steadily risen since then, with a major increase since the mid-1980s which was accompanied with a change in their recognition among economists (Durham et al., 2007). Among others, Falk and Heckman (2009) nicely demonstrate this development by the following quotes of Samuelson and Nordhaus. In the 1985 edition of their book *Principles of Economics*, the authors write: "Economists (unfortunately) ... cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe." Seven years later in the 1992 edition, this view has changed and Samuelson and Nordhaus acknowledge the virtue of controlled experiments also in economics: "Experimental

economics is an 'exiting new development'."

The key advantage of laboratory experiments is the controlled variation of one variable keeping the other conditions fixed. This guarantees insights into the behaviour of individual economic agents which are difficult to obtain using conventional econometric techniques. Empirical data always include a large variety of environmental factors and a disentanglement of their influences is difficult if not impossible. In a laboratory experiment, these factors can be controlled allowing research to identify and test causal relationships (Falk and Heckman, 2009; Levitt and List, 2007). Lab experiments have proved to be particularly useful when testing a specific theoretical model. One can test predictions made by the theory because real agents play the exact game with real monetary rewards in the lab. Additionally, it is relatively easy to test competing explanations for the observed behaviour in case the model is rejected (Charness and Kuhn, 2011). Another strength of lab experiments is to provide behavioural insights when the theory makes no clear prediction as in games with multiple equilibria. In such cases, they generate information on people's actual behaviour where the existing model is no guide to what should happen (Charness and Kuhn, 2011).

A critical underlying assumption is that the results generated in the artificial environment of laboratory experiments can be generalized (external validity). That is, they can be considered valid in the broader environment (Levitt and List, 2007). Most this concerning objections including the fact that participants are usually undergraduate students, that stakes are low or that the environment is too abstract could be refuted by field experiments (see for example Charness and Kuhn (2011), Falk and Heckman (2009) or Cooper and Kagel (2009)). Nevertheless, Holt (2006) points out that caution is advised especially when social

context is critically important.

While the laboratory only provides a model of the field environment missing many details that may influence agents' actions, Charness and Kuhn (2011) conclude that lab experiments nevertheless provide useful qualitative insights. In how far quantitative levels of behaviour apply to naturally occurring settings, should however carefully be considered. Lab experiments and field studies therefore should be seen as complementary.

Chapter 3

Literature Review

3.1 Empirical evidence of other-regarding behaviour

"[E]xperimental economists have gathered overwhelming evidence that systematically refutes the self-interest hypothesis and suggests that a substantial fraction of the people exhibit social preferences, in particular, preferences for reciprocal fairness."(Fehr and Fischbacher, 2002, p. C1)

One of the first experiments that showed the relevance of other-regarding behaviour was the ultimatum game published by Güth et al. (1982). In this very simple two-player game, the first player can make a proposal on how to split a certain amount of money. The second player can accept or reject the proposed division. If he accepts, both players receive an amount according to the first player's suggested partition. If he rejects, both players earn nothing. While there exist several Nash equilibria, the only subgame-perfect equilibrium under standard "selfish" assumptions predicts that the second player accepts any positive amount of money, and – anticipating this reaction – the first player allocates the smallest

possible amount ϵ to the second mover and keeps the rest for himself. In reality, however, experimental evidence robustly shows that the large majority offers the second player between 40 and 50 percent of the available money. Moreover, 40 to 60 percent of second players reject proposals in which they receive less than 20 percent of the available surplus (Fehr and Schmidt, 2006; Cooper and Kagel, 2009). Slonim and Roth (1998) further showed that neither low stakes nor missing learning opportunities are the ultimate reason of the observed behaviour, thereby refuting the claim inter alia made by Binmore et al. (1985) that findings mirror erroneous anomalies.

In fact, even if any strategic uncertainties for the first movers were removed, pro-social behaviour did not vanish. The latter was shown using the dictator game which modifies the ultimatum game in such that the second mover's role is reduced to a passive acceptor: The first mover's division proposition is immediately paid out so that the first mover does not bear any risk of rejection. In contrast to the classical predictions, the first mover usually does not keep the entire surplus for himself but assigns positive amounts up to 50 percent of the pie to the other person (e.g. Forsythe et al., 1994).[1]

An interesting companion game is the slightly more complex investment (or trust) game (Berg et al., 1995), which again consists of two players and two stages. Both agents are initially endowed with a certain amount of money e. The first mover has the opportunity to send any integer amount s between zero and e to the second mover, who receives ks (typically $k = 3$). The second mover can then return an amount r between zero and ks. The investment game resembles a typical exchange

[1]The average amount allocated to the second mover lies between 10 and 25 percent with modal allocations at 50 percent and zero (Fehr and Schmidt, 2006).

situation in which the first mover gives up a sure payoff for an anticipated future benefit. But receiving this future benefit is contingent on the second mover's behaviour. It is therefore in principle a dictator game in which the second mover dictates the final payoff allocation with the difference that the first mover determines with his sent amount the surplus that is being shared. Again, under classical assumptions, a self-interested second mover will never return any money, i.e. $r = 0$. Anticipatory first movers should therefore transfer nothing, i.e. $s = 0$. In experiments, positive amounts of money are typically sent and returned albeit there are considerable individual differences. First movers invest on average about half the maximum and second movers repay on average approximately s. The most interesting observation that can be made is that the amount second movers return increases on average with the initial transfer s if the change in s is sufficiently high (Fehr and Schmidt, 2006).

The investment game is a specific kind of trust game. The broader category of trust situations can be characterized through a sequential process where the trustor moves first and the trustee observes the trustor's behaviour before making his own decision. Bacharach et al. (2007) characterizes the most basic form as a two-person game, whose extensive form is shown in Figure 3.1 and whose parameters fulfil the following criteria:

$$\begin{aligned} b < a \qquad & \text{Exposure,} \\ a < c \qquad & \text{Improvement,} \\ z < y \qquad & \text{Temptation.} \end{aligned}$$

Both players have two strategies to choose from. The trustor can either *trust* or *withhold*, and the trustee can decide between $fulfil$ and *violate*. The inequalities imply thereby that a) the trustor accepts a risk

by trusting as he is worse off if the trustee violates, b) the trustor can be better off by trusting (i.e. if the trustee fulfils), and c) the trustee has a monetary incentive to violate the trust. While, according to Bacharach et al. (2007), most writers on trust agree that these inequalities are fundamental, some make more restrictive requirements on the parameters (e.g. Bacharach and Gambetta, 2001). Particularly interesting is the additional inequality of $x < z$. This "mutual gain" condition states that the pair $trust/fulfil$ results in a Pareto-improvement compared to the status quo. However, we often think of trusting acts even if this condition is not satisfied: We frequently rely on another person to behave in a way that makes him worse of, e.g. trusting the neighbour to water the plants. Under standard assumptions of purely self-interested agents, the prediction of any trust game is $(withhold, violate)$: Foreseeing that the trustee will play the weakly dominant strategy $violate$, the trustor plays $withhold$ and there is no trusting. Yet, not only the previously discussed investment game but many other laboratory experiments show that the standard prediction is systematically violated (see among others Fehr and Gächter, 1998; Dufwenberg and Gneezy, 2000; Schotter and Sopher, 2006). In all studies, 50 percent or more of subjects in the role of the trustor play $trust$ and many subjects in the role of the trustee $fulfil$.[2]

Additionally, a great many of various different games can be found that provide more evidence of social behaviour. They will be neglected here as they are not particularly relevant for this thesis. Modifications of the previously described games will be discussed in the sections where they are of interest.[3]

[2]In fractional versions as the investment game, trustors give half or more of their endowment and trustees fulfil to a substantial degree (return money).

[3]Note that there also exists a strand of literature investigating envy (Bolton, 1991), vendettas (Bolle et al., 2014) or similar destructive behaviour (e.g. Nikiforakis,

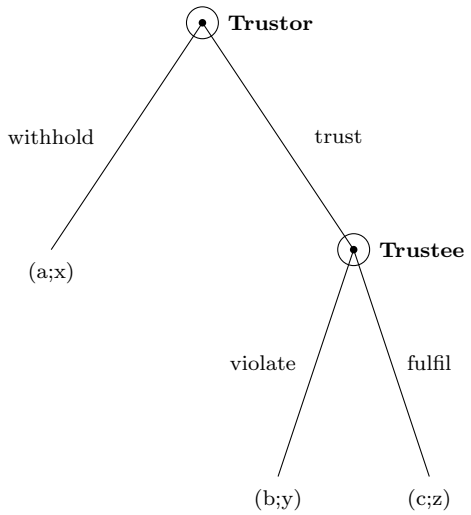


Figure 3.1: Extensive form of a typical trust game.

3.2 Theoretical approaches to pro-social behaviour

The abundance and prominence of fairness behaviour has given rise to a growing literature on models of "social" or "other-regarding preferences" that assume that people's well-being is not only determined by their own material payoff. Instead, these models try to explain when and how an agent takes others' monetary gain into account. All agents are assumed to behave rationally so that the well known concepts of utility maximization and game theory can be applied to analyse agents' optimal behaviour (Fehr and Schmidt, 2006). However, in the theories of other-regarding preferences, arguments beyond material self-interest enter an agent's utility function. Typical examples for such arguments are others' (material) well-being (distributional preference models) or functions thereof. For the latter, people's weight put on the others' well-being may depend on others' intentions or others' observed behaviour (intention-based or

2008).

reciprocity models), others' payoff expectations (guilt aversion models) or others' other-regarding concerns (type-based models). Below, I will shortly discuss the first three classes – models of distributional, models of intention-based other-regarding preferences and models of guilt aversion – in a two-player context. The list of discussed models is most certainly not exhaustive but will provide an overview over the existing theories.[4]

Additionally, I will shortly discuss one more model of social norms. While other-regarding preferences are often seen as the stabilizing factor behind the coherence and enforcement of social norms (and Fehr and Schmidt (2006) even claim that cooperative institutions enforcing rules and norms may only exist because of the existence of other-regarding preferences), others such as Krupka and Weber (2013) argue the other way around by stating that measuring other-regarding preferences is economists' indirect way of measuring social norms. They propose a model of norm-compliance where people do not care about others' payoff per se but people care about behaving in a manner consistent with social norms. The model by Krupka and Weber (2013) and the empirical implications thereof will be discussed at the end of this section.

3.2.1 Distributional social preferences

Motivated inter alia by results from the dictator and the ultimatum game, distributional models of social preferences were developed in order to account for the found results. These initially suggested models focus on an agent's actions that are motivated by the properties of the outcome. In a two-player context, models of this strand assume that the utility of a certain payoff of a game depends only on player's own income x_i and

[4]For a detailed literature review, see Fehr and Schmidt (2006).

the other player's payoff x_j in the final outcome. The simplest altruism model by Andreoni and Miller (2002) states that an agent is altruistic *iff* his utility is increasing in own and other's income, i.e. $\frac{\partial u(x_i,x_j)}{\partial x_i} > 0$ and $\frac{\partial u(x_i,x_j)}{\partial x_j} > 0$. Thus, the indifference curves have the "typical" negative slope indicating that own as well as other's payoff is always a "good". To fit their data, they suggest the most familiar form of the constant elasticity of substitution (CES) utility functions:

$$u(x_i, x_j) = \left(x_i^\alpha + \theta x_j^\alpha\right)^{\frac{1}{\alpha}},$$

where $\alpha \in (-\infty, 1]\backslash\{0\}$ and $\theta \geq 0$.

Cox and Sadiraj (2007) propose a modified CES utility function in their model of egocentric altruism. Their model includes non-linear indifference curves resulting from the utility function with the parameter restriction $\theta \in [0, 1)$:

$$u(x_i, x_j) = \begin{cases} \left(x_i^\alpha + \theta x_j^\alpha\right)\alpha^{-1} & \alpha \in (-\infty, 0) \cup (0, 1], \\ x_i x_j^\theta & \alpha = 0. \end{cases}$$

The major advantage of this modification is the possible incorporation of reciprocity concerns (which I will discuss in Section 3.2.2) as the altruism coefficient θ does not necessarily need to be restricted to be positive.

These altruism models already include a concern for the relative standing between the agents which is captured by the convexity parameter α. In the following models of distributional preferences, the importance of this relative standing is increased as they assume a qualitative change in preferences at equality. The probably most prominent models postulate that people are "inequality-averse" (Fehr and Schmidt, 1999; Bolton and

Ockenfels, 2000): They do not like their payoff to differ from the equal share (and they especially do not like to fall behind). While an agent likes money, he also compares his own payoff with the payoff of the other(s) and experiences distress from compassion if he is better off and from envy if he is worse off. The two-player version of the Fehr and Schmidt (1999) model has piecewise linear indifference curves for difference-averse preferences over the income for himself x_i and the other's income x_j. The corresponding utility function is linear in own earnings and the difference in earnings:

$$u(x_i, x_j) = x_i - \alpha \cdot max[x_j - x_i, 0] - \beta \cdot max[x_i - x_j, 0]$$

or

$$u(x_i, x_j) = (1 + \alpha)x_i - \alpha x_j \quad \text{for} \quad x_j > x_i,$$

$$u(x_i, x_j) = (1 - \beta)x_i + \beta x_j \quad \text{for} \quad x_j \leq x_i.$$

The marginal rate of substitution parameters are assumed to satisfy $0 \leq \beta \leq \alpha$ and $\beta < 1$. The latter assures that one always likes own income while the former assumes that the marginal rate of substitution between own and other's income depends on who has the higher income whereby an agent suffers more from inequality if he falls behind in payoffs than if the other earns less.

The Bolton-Ockenfels utility function only differs in that it takes a non-linear form: $u(x_i, x_j) = v(x_i, x_i/(x_i + x_j))$, where v is a non-decreasing function which is concave in the first argument and strictly concave in the second argument, the relative income, with a maximum at 1/2. The second assumption on v states that an agent's well-being increases in other's income if better off, but decreases in other's income

if behind (holding own material payoff constant).

While these models are good in explaining positive acts to others such as the generosity observed in dictator games or the positive back-transfers in investment games, they are less equipped to describe negative acts as the rejections in ultimatum games. Additionally, the underlying assumption that the agent's social preferences only depend on the final distribution of payoffs often seems too restrictive. Only the intrinsic properties of outcomes are assumed to be decisive while alternative choices that players face or the choice process are irrelevant. As a consequence, other (non-distributional) models of social preferences that emphasize the role of intentions and/or the way a decision situation came to place have been developed.

3.2.2 Intention-based social preferences

The second approach of other-regarding preferences tries to explain predictions inconsistent with self-regarding preferences by the agent's desire to react to another's intention. Agents' behaviour is no longer solely motivated by the final outcome but also by the way this outcome has been achieved. People pay attention to the perceived intentions that drive the other players' actions and may be willing to reward and/or punish certain types of behaviour. Individuals' preferences can thus become more or less altruistic depending on the perceived intention of another subject.

The idea of an attribution of intentional states to others originates from what cognitive scientists call "mindreading" or "folk psychology" (Baron-Cohen, 1997). Humans routinely attribute mental states such as beliefs, deliberateness or desires to others in order to explain their

actions. Similarly, we use this attribution of mental states to constantly and mostly unconsciously predict other's behaviour (Baron-Cohen, 1997; McCabe et al., 2003).

Theories of this strand typically focus on reciprocity, i.e. subjects' responsiveness to behaviour perceived as nice with positive reciprocal (more friendly) actions and to as unkind perceived actions with negative reciprocity (more nasty behaviour). Hence, preferences depend not only on material payoffs but also on the player's interpretation of his opponent's behaviour, and accordingly his belief about the reason for a chosen action, i.e. the intention. A key question for reciprocity models is how individuals assess the kindness of a particular action.

Blount (1995) shows as one of the first that the intentional act, the free will, behind a choice is relevant for its perceived kindness. She shows that rejection rates are much lower in the ultimatum game when low proposals were generated by a computer and not by another participant. This indicates that fairness norms are only or at least more often imposed if payoff allocations are reached deliberately through choices of the opponent. Similar results are found by Offerman (2002) and Falk et al. (2008).

An additional way of evaluating kindness is the comparison of a choice with the alternative acts that could have been chosen. Evidence for the relevance of the available actions, the strategy space, is shown for example by Falk et al. (2003): Second movers in the ultimatum game rejected the same offer less often when it is the most generous offer than when it is the least generous in the first movers opportunity set.

The models by Rabin (1993), Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006) go one step further. They propose that for the evaluation of another person's kindness, not only the alternatives are

relevant but also the beliefs about why the other chose a particular action. To evaluate the intention behind the act, one therefore needs to form beliefs about what the other person believes oneself to do. Their models are based in the literature of psychological game theory (Geanakoplos et al., 1989), where players' preferences do not only depend on material payoffs but also on a player's beliefs about other's choices or beliefs. In particular, perceived kindness in zero-sum games depends crucially on a player's second-order belief: The more the other hopes/expects to accrue for himself from the final payoff allocation, the less kind the other's choice is. Hence, given a certain choice, the higher one's second-order belief, the less kind the other's choice is interpreted and in turn the less kind one's own action.

These models provide quite sophisticated theories of reciprocity but unfortunately, they are not very tractable even in simple games and often yield many equilibria. To avoid these problems, Cox et al. (2007) and Cox et al. (2008a) propose a more modest reciprocity or "revealed altruism" approach that relies solely on observable actions but which is a pure preference and not an equilibrium model. Their model builds on the egocentric altruism model of Cox and Sadiraj (2007, cf. Section 3.2.1) but modifies it in such that the "emotional state" θ that determines the weight put on the other agent's (material) well-being is no longer a parameter but a function of the kindness or unkindness of the other's observed choices. For this purpose they define a "reciprocity" or "revealed kindness" variable r and assume that θ is an increasing function of r.[5] The revealed kindness r, in turn, is an increasing function of the "generosity" of the other's choice. It is defined as: $r(a_j) = x_i^{max}(a_j) - x_i(a^0)$. The

[5]In their paper, Cox et al. (2007) also include status as a positive influence of θ which will be neglected in the further analysis as it is irrelevant for this thesis.

revealed kindness of a first mover's action a_j is the difference between the maximum payoff a second mover can guarantee himself given the choice a_j, $x_i^{max}(a_j)$, and his payoff in a reference action a^0 that is "neutral in some appropriate sense". By replacing the restriction of $\theta \in [0, 1)$ by $\theta < 1$, Cox et al. (2007) explicitly allow for negative concerns for the other's payoff. The essence is that the first mover's kindness is *revealed* by the generosity of the opportunity set that his choice induces for the second mover, which is evaluated by a comparison to the other options that were available to the first mover.

3.2.3 Guilt aversion

Charness and Dufwenberg (2006) argue that not only the kindness of an action determines reciprocity but that people may help the other because they would feel guilty if they did not.[6] This idea builds on psychological research where guilt is considered a form of emotional distress which is based in social relationships. Typically, guilt is generated when behaviour has deviated from social standards that have been internalized by the individual; its anticipation motivates and encourages pro-social behaviour (Lazarus, 1991). Guilt in the context of social preferences refers to a guilt-feeling as the consequence of not living up to others' expectations (Baumeister et al., 1994; Battigalli and Dufwenberg, 2007). In what follows I focus on the theory of "simple guilt" proposed by Charness and Dufwenberg (2006) and extended by Battigalli and Dufwenberg (2007). It postulates that players experience a utility loss if they believe that they let others' payoff expectation down.[7] The basic idea is

[6]The focus lies only on positive reciprocity; punishing behaviour cannot be accommodated.

[7]In contrast, the theory of "guilt from blame" (Battigalli and Dufwenberg, 2007, 2009) postulates that players experience a utility loss if they believe that *others believe*

that player i suffers from guilt to the extent that he believes that player $j \neq i$ gets a lower monetary payoff than i believes j expects to receive. Thus, a player's utility depends directly on beliefs. Using psychological game theory (Geanakoplos et al., 1989), an agent i's utility function in a two-player game can be defined as:

$$u(z, \alpha_{ji}) = x_i(z) - \theta max\{0, E_{\alpha_{ji}}[x_j] - x_j(z)\},$$

where z is the outcome of the game (the reached terminal node), $x_i(z)$ ($x_j(z)$) is the monetary payoff of agent i (j) at z, α_{ji} is player j's ex-ante belief on i's play of the game, $E_{\alpha_{ji}}[x_j]$ is j's subjective expected payoff calculated using α_{ji}, and θ is an exogenously given positive constant. While player i's utility is strictly increasing in own payoff, he also experiences a psychological cost of not fulfilling player j's ex-ante payoff expectations. The extent of the latter is given by θ which measures i's guilt-sensitivity. Player i feels guilty about disappointing player j and will feel guiltier the larger the difference between between $E_{\alpha_{ji}}[x_j]$ and x_j. A guilt-averse agent will consequently assign more money to agent j when he expects more. There is no gain from exceeding the other's payoff expectations, though. Note further that player i has to form and use his second-order beliefs in order to compute his expected utility from his different action choices since he typically does not know player j's first-order belief α_{ji}.

In order to test their model, Charness and Dufwenberg (2006) conducted several trust games with a preceding promise stage: Before the actual trust game started, the second mover could promise the first mover to reciprocate if the first mover trusts him. They found that communi-

that they believe that they let others' payoff expectation down.

cation (i.e. promising) does not only enhance trust, reciprocation and efficiency but more importantly they found support for guilt aversion: A second mover was significantly more likely to help the first mover the more he believed that "his" first mover expected him to help.[8]

3.2.4 Social norms

Another approach to explain social behaviour is based on people's desire to comply with social norms. Social norms typically refer to a mutual understanding among members of a group regarding a certain behaviour or action, rather than outcomes. Each individual in a particular group shares its judgement regarding the (in)appropriateness of behaviour with the whole group and this common judgement is a social norm (see, for instance, Krupka et al., 2011; Young, 1998; Burke and Young, 2010). The force of social norms thereby stems from people's willingness to punish (or reward) others' deviation from (or adherence to) social norms within a population on the one hand, and from the experience of positive or negative emotions produced by one's own adherence or deviation from a social norm on the other hand (Elster, 1989; Fehr and Gächter, 2000).

The importance of social norms on people's behaviour has been recognized for a long time in psychology or sociology. In economics, contrariwise, social influences have more or less been ignored (Krupka and Weber, 2013). In fact, the founders of economics such as Mill acknowledged them more than neoclassical theorists of the last century (Burke and Young, 2010). Furthermore, social norms often served mainly as a post-hoc interpretation of otherwise unexplained phenomena (Fehr and Gächter, 2000; Con, 2003; Ostrom, 2000). It is only until very recently

[8]For a more detailed discussion on the empirical evidence on guilt aversion, please refer to Section 6.2.

that researchers started to incorporate norms into economic models, investigating how they affect people's behaviour. This development was mainly driven by the experimental evidence of systematically deviating behaviour from game-theoretic predictions based on self-interest. This evidence on deviations was interpreted as the impact of social preferences (how players feel when others earn more or less money) and social norms (what players expect and feel obligated to do) (Camerer and Fehr, 2004). However, often the distinction between both is not made or unclear. Krupka et al. (2011) and Krupka and Weber (2013) go as far as claiming that measuring other-regarding preferences is economists' indirect way of measuring social norms, i.e. the norm of fairness or inequity aversion (e.g. Fehr and Falk, 1999; Fehr and Gächter, 2000), of reciprocity (e.g. Dufwenberg and Kirchsteiger, 2004), or the norm to honour obligations (e.g. Hart and Moore, 2008).

More recently, Krupka and Weber (2013) introduced a model of social norms and maybe even more importantly a method on how to identify and measure social norms. Following the literature, they define social norms as "collective perceptions, among members of a population, regarding the appropriateness of different behaviours. They are things that people in the population jointly recognize one should or should not do, and people who belong to the population expect others to be aware of and understand this agreement." In their model, a social norm $N(a_k) \in [-1, 1]$ is a collective judgement that assigns a degree of (in)appropriateness to every action a_k available to the decision-maker (and which can be empirically determined). This definition allows actions to vary in the degree to which they are perceived as socially (in)acceptable rather than only being right or wrong. The crucial assumption is that people's utility does not only depend on the money they

obtain but also on the degree to which their actions reflect social norms. This means that decision-makers have a preference to take actions that are seen as socially appropriate and to avoid actions considered as socially inadequate. Translated in a utility framework, these consideration read:

$$u(a_k) = V(\pi(a_k)) + \gamma N(a_k),$$

where $V(\pi(a_k))$ is the value an individual puts on his monetary payoff $\pi(a_k)$ resulting from his action a_k and $\gamma \geq 0$ represents the degree to which a decision-maker cares about complying with social norms.

It follows that individuals with a preference for norm-compliance (i.e. $\gamma > 0$) may choose different actions (select differing payoff pairs) across choice environments even if the sets of available actions and payoffs are exactly the same, but the two environments differ in their social norms.

Based on the introduction of this model (and the associated elicitation method for social norms), a horse-race between the two (different?) explanations of pro-social behaviour has been started by asking the question: Is pro-social behaviour driven by stable other-regarding preferences or rather by the compliance with social norms (i.e. socially appropriate behaviour in a given context). Krupka et al. (2011) find that elicited social norms have a substantial explanatory power in dictator and Bertrand games. Krupka and Weber (2013) furthermore show that context-dependent variations in the dictator game, which cannot be captured by the models of social preferences, can be explained by social norm compliance. These findings are in contradiction with those of Gächter et al. (2013) who ask the same question for a different context, examining peer effects in a three-person gift-exchange game. Their results

suggest the superiority of distributional social preferences to preferences for behaving socially appropriate.

Part II

Three Essays on Intention-based Social Preferences

Chapter 4

Driving a hard bargain is a balancing act: The importance of reciprocity in bargaining

Abstract[1]

We investigate the effect of opening offers in bargaining. We find that, even if a first offer is costless to reject, it has a significant impact on the bargaining outcome. Opening offers convey information on the player's reservation value induced by his social preferences and they are most often accepted when they are not above the equal split. However, offers which request much more than the equal split induce punishing counter-offers triggered by the responder's social preferences. The bargaining outcome is therefore critically influenced by the balance of toughness and kindness signalled through the offers made during the haggling phase.

[1]This is work co-authored by Lionel Page.

> *The usual haggling process is based on imperfect information, the hagglers trying to propagandise each other into misconceptions of the utilities involved.* Nash (1953), Two-person cooperative games
>
> *Lest readers think erroneously that it's always wise to bargain tough.* Raiffa (1982), The Art and Science of Negotiation

4.1 Introduction

Bargaining is pervading in economic and social interactions and it has been a natural object of study for economists. The large economic literature on economic bargaining has brought many insights about how bargaining outcomes are determined depending on the bargaining power, exit options and preferences of the bargainers. However, still little is known about the negotiation process itself: What is a good opening proposal, how long should you stick to a proposal and how much should you change your proposal when it is a deadlock? Negotiation in real-economic situations often seems an art which requires expert practice to excel. Studying the negotiation process in the field is unfortunately challenging since negotiations are typically characterized by imperfect information and the use of messages which are hard to measure and quantify.

This paper investigates a critical aspect of negotiations: the "haggling" process where players exchange split proposals which are costless to reject. We focus here on the effect of first proposals. To do so, we design a game where the zone of possible agreements is known (similarly to an ultimatum game). Under standard assumptions of common knowledge of rationality and payoff-maximization, first proposals should

not influence the final outcome. However, the fact that bargainers may have social preferences transforms this game into a game with imperfect information where players' preferences are not common knowledge. Players can use the haggling process to try to influence each other's beliefs about personal preferences (Nash, 1953). To assess the effect of first proposals on the bargaining process, we study the effect of a wide range of first proposals on the reaction of the player receiving it. We elicit both the actions of the receiver and his first- and second-order beliefs. This design allows us to study whether and how the level of the opening proposal influences bargainers' beliefs, their actions and the final bargaining outcome.

This study contributes to several important strands of literature. First, it extends the literature on the role of social preferences in bargaining by investigating how social preferences do not only limit the range of acceptable outcomes, but also constrain the negotiation process itself. Bargaining experiments have shown that bargaining outcomes are influenced by players' preferences over payoff distributions (Camerer, 2003a), and by their preferences over the intentions of other players (Blount, 1995; Offerman, 2002). We complement these results by studying whether social preferences can be triggered during the sequence of offers and counteroffers in a negotiation and consequently influence the final bargaining outcome. In the haggling process, players making an offer can aim to signal a preference for fairness suggesting that unfavourable splits will be rejected; they can also try to bargain tough to secure the best outcome possible. The intention-based preferences of the players receiving these offers can play a role if proposals are perceived as signalling something about the likely intention of the other player. We investigate these possibilities by studying carefully the effect of a first proposal in a nego-

tiation game.

Second, this paper adds to the literature on bargaining with reputation where a player has the possibility to build a reputation for stubbornness by sending an initial message to the other player (Abreu and Gul, 2000; Wolitzky, 2012; Embrey et al., 2014). Opening proposals in real-world negotiations are often intended to signal the toughness of one's bargaining strategy. Yet, little is known about whether the first proposal has indeed an impact on the other bargainer's beliefs.[2] We elicit the belief of the player receiving the first proposal (Responder) about the minimal amount the player making the proposal (Proposer) would accept. Doing so, we can measure whether first proposals convey any information about the player's final bargaining stance.

Third, our study complements the research on the role of communication in bargaining games. Experimental studies have found that "cheap talk" phases before the bargaining itself influence players' strategies and therefore the bargaining outcome (Croson et al., 2003; Rankin, 2003; Anbarci et al., 2015). Moreover, proposals themselves can be used to communicate feelings and intentions to the other player (Xiao and Houser, 2005). In the context of our lab experiment, we isolate and study a simple and precise piece of information: the level of an opening offer which is costless to reject. Because it is costless to reject is does not formally affect the bargaining power of the player making the proposal. Whatever role this proposal has on the bargaining process, it has to come from its role as a communication tool.

Our results are striking in what they reveal about the bargaining pro-

[2] An interesting related study by Goldreich and Pomorski (2011) looks at the effect of the decision to initiate the bargaining process by making a first proposal. In our case, the initiating role is pre-assigned and we study the effect of the level of the first proposal.

cess. First, we find that the first proposal has a substantial effect on the bargaining outcome even though it is costless to reject. Proposers' first proposals are correlated with their minimal acceptable amounts and therefore carry some information about the Proposer's likely refusal of unfavourable splits. We also find that once the first proposal has been made, our Proposers are credibly obstinate as they tend to reject unfavourable counter-proposals with a high probability. In the end, our Proposers are able to get a high proportion of the pie to be divided even though, under standard assumptions, they have as much bargaining power as in the ultimatum game – which is none.

Second, we find that the Responder's intention-based social preferences are triggered by first proposals. Proposals which favour the Proposer are not only rejected but often lead to low counter-offers from the Responder. In a substantial number of cases, the Responder chooses a punishing counter-offer which is lower than what he believes to be the Proposer's minimal acceptable amount. Looking at the mechanisms behind these intention-based preferences, we are able to investigate different theoretical explanations. Using the elicited first- and second-order beliefs from the players, we do not find evidence suggesting that these beliefs drive players' behaviour as suggested by psychological game-theoretic models of reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004). We however find support for the Levine model of reciprocity and spitefulness where agents care about the type of the other player (Levine, 1998). We also find evidence suggesting that players react negatively to proposals which are perceived as disrespectful as suggested by Yamagishi et al. (2012) in line with recent research pointing to individual preferences for self-esteem (Bénabou and Tirole, 2006; Ellingsen and Johannesson, 2008), the demand for respect (Eriksson and Villeval, 2012)

and status (Besley and Ghatak, 2008; Heffetz and Frank, 2008; Charness et al., 2010) in social interactions.

The remainder of the paper is organized as follows: The next section ties our research in the context of the existing literature. Section 4.3 introduces our experimental design and outlines our research hypotheses. It is followed by the analysis of our data and the obtained results in Section 4.4. Section 4.5 concludes with a short summary and the discussion of our findings.

4.2 Related literature

Bargaining experiments have found systematic departures from game-theoretic predictions. In the case of alternating bargaining games (Ståhl, 1972; Rubinstein, 1982), the subgame-perfect equilibrium predicts for similarly impatient and purely self-interested bargainers who have complete information an immediate agreement on a bigger share for the party that makes the initial offer. The evidence from bargaining experiments however reveals that participants do not play rationally in order to maximise solely their own payoff. As a consequence, the literature in behavioural game theory has proposed that bargainers may have some behavioural types whose strategies differ from standard assumption.

In particular, the experimental evidence shows that bargainers care about the "fairness" of the bargaining outcome (Roth, 1995b). The robustness of this evidence has contributed to the motivation of models of social preferences where players care about others' payoffs and how these compare to their own payoff (e.g. Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002). Besides distributive concerns, it has been suggested that players also care about each other's

intentions and that players react to the perceived kindness of other players. Blount (1995) and Offerman (2002) showed that when the first mover's "action" is determined by a random device, responders are less likely to reject smaller offers than when the offer is made intentionally by the first mover. More evidence for the importance of intentions on rejections can be found in Falk et al. (2003). They show that a (8;2) split in favour of the proposer is more likely to be accepted if the only other alternative was even more unfair (10;0) than if the only other alternative was equitable (5;5).

Models of intention-based preferences have been developed to rationalise these observations. First, in the framework of psychological game theory, Rabin (1993) and Dufwenberg and Kirchsteiger (2004) have proposed models of reciprocal preferences where players deduce the other players' kindness from their beliefs about the other players' intended actions. In such a framework, a player A will respond kindly to a move made by a player B when A believes that B made this move with the intention that his move would lead to a good (in some sense) outcome for the player A. Second, another influential model by Levine (1998) suggests that people react the the type of the other player. An altruist player may want to act altruistically when facing other altruists but spitefully with other players who are no altruists. Finally, rejections of low offers in the ultimatum game have also be explained by a concern for respect. Rejections can be driven by "wounded pride" (Straub and Murnighan, 1995): Disadvantageous offers are seen as signalling a lack of respect on part of the other player. This explanation has received support from the observation that the rejection decisions in the ultimatum game are not correlated with reciprocal attitudes in other distributive games (Yamagishi et al., 2012). Instead, such a behaviour could be driven by a concern

for maintaining a reputation as a tough bargainer, which could have been selected by evolution (Burnham, 2007; Embrey et al., 2014).

If bargainers have different types, it potentially increases the importance of communication in order to find an agreement in a negotiation. Bargainers may want to communicate their true type or lie about it. Real-world bargaining situations typically allow players to communicate either verbally or through (costless) proposals and counter-proposals. This communication phase usually contains declarations about desired outcomes and reservation values which are not verifiable. From a non-cooperative game-theoretic point of view with self-centred bargainers, such communication is "cheap talk" (Crawford and Sobel, 1982) and should therefore not be expected to impact the final outcome of the bargaining process.[3] However empirical evidence shows that cheap talk in bargaining games can have an effect on the outcomes by influencing players' beliefs (Croson et al., 2003; Rankin, 2003; Tingley and Walter, 2011; Kriss et al., 2013; Anbarci et al., 2015).[4] Players' claims in pre-play communication seem to be interpreted as if they contain an element of truth. Models of bargaining with reputation (Abreu and Gul, 2000; Wolitzky, 2012; Embrey et al., 2014) show that such signals should indeed have an effect if the population contains bargainers with "obstinate" behavioural types who refuse low proposals.

The effective communication between bargainers is not necessarily re-

[3]This is the case because communication has no direct payoff implications. Players with competing interests, as in a bargaining process, have no incentive to communicate truthfully. Theoretically, cheap talk should consequently not have any effect on the equilibrium strategies of the players in most games.

[4]Early experimental studies with unstructured bargaining designs already indicated that free communication between bargainers plays a role. In particular, face-to-face bargaining has been found to improve the likelihood of a bargaining agreement, possibly because it provides a richer array of channels of communication "including tone of voice, body language and facial expression" (Roth, 1995, p. 296).

stricted to explicit communication signals. Proposals themselves can be used to communicate intent and/or emotion to the other player. Studies have found that allowing players to communicate with messages reduces the proportion of rejections in the ultimatum game (Xiao and Houser, 2005; Andersson et al., 2010). The rejection of a proposal seems to be used to convey a message of dissatisfaction. When the dissatisfaction can be either attenuated or expressed by explicit communication between players, the rejection of a proposal loses its appeal as its role as a signal is mitigated or even redundant. And vice versa, if no cheaper communication channel is available, the rejection is used to signal one's unease with a certain proposal.

If proposals can send a signal, the players can potentially use the sequence of proposals and counter-proposals to try to convey the impression that they have a high reservation value in order to claim a larger slice of the pie. Conversely, they may want to convey a signal of agreeableness to increase the chance of a deal. If communication has an effect on players' perception of each other's intention, it can play a substantial role in the determination of the bargaining outcome because of bargainers' intention-based preferences. And if proposals convey messages of intent, then bargainers have to consider the message they may be sending when choosing an opening proposal. Whilst we know little about the mechanisms behind real-world negotiation strategies (for an exception, see Goldreich and Pomorski, 2011), the art of negotiation is part of the training in Business Schools. And it indeed stresses the importance of making reasonable first proposals when initiating a bargaining process (Raiffa et al., 2002).

4.3 Experiment

4.3.1 Experimental design

We designed a two-stage alternating-offer bargaining game with no shrinkage of the pie from the first to the second period (see Figure 4.1). In a first stage, the Proposer makes a proposal on how to divide a pie of \$10 between himself ($\mathcal{P}_P$) and the Responder ($\$10 - \mathcal{P}_P$).[5] The Responder can then either accept or reject the proposal. If the Responder accepts, both players receive the amount corresponding to the Proposer's suggested partition. If he rejects the proposal, the Responder makes a counter-proposal to the Proposer regarding the split of the money ($\mathcal{P}_{CP}, \$10 - \mathcal{P}_{CP}$). The Responder's counter-proposal is then either accepted or rejected by the Proposer. If the Proposer accepts it, the suggested partition is implemented. If the Proposer rejects the counter-proposal, both players earn nothing. For simplicity, we characterize the proposals and counter-proposals in the remainder by the amount they ascribe to the Proposer in the suggested split of the pie.[6]

The absence of shrinkage makes the message (in the form of a split proposition) in the first stage costless to reject. Under standard assumptions, in particular for money being the only carrier of utility for the bargainers, all the bargaining power relies in the hand of the Responder who can make the final take-it or leave-it proposition. In the subgame-perfect equilibrium of the game, the Responder refuses any initial proposal which gives a positive amount to the Proposer and proposes zero in his counter-proposal which is accepted by the Proposer. In that sense,

[5] All monetary amounts are in Australian dollars.

[6] In the experiment, the wording was neutrally for both players to characterize a proposed split between oneself and the other.

the Responder is in a position very similar to a proposer in the ultimatum game.[7] The existing experimental research in behavioural game theory shows that bargainers do not care only for their own money. They also care about the fairness of the split. As a consequence their reservation value is typically higher than zero. Our haggling game is therefore characterized by imperfect information about the players' preferences (e.g. minimal acceptable proposals). Proposers can use the initial proposal to try to influence the Responder's belief about their preferences.

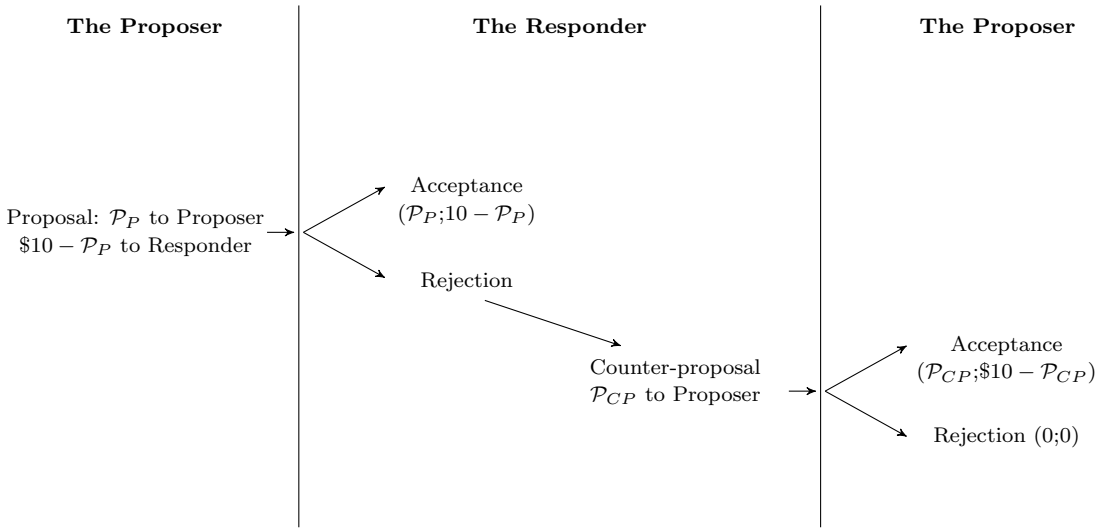


Figure 4.1: Game tree of our two-stage alternating bargaining game without shrinkage of the pie.

The experiment took place between March and September 2014. In total, we collected data on choices of 158 subjects in total in the bar-

[7]Note that while the first proposal is costless to reject, it can also be accepted. For that reason, it is not a cheap-talk message but bears a cost for the Proposer: For any proposal assigning him less than the full amount, he accepts the possibility to forego higher amounts. An ultimatum game with a pure cheap-talk request in a pre-play communication phase was investigated by Rankin (2003). The study unexpectedly found a negative effect of the option to make a request. In light of our results, we interpret this finding as the consequence of the high level of requests observed in the study (and possibly from the emotional effect of the words "I request" used by the receiver).

gaining experiment. Each participant played either the Proposer or the Responder (79 participants in each each role). There were ten experimental sessions with ten to 20 subjects. It was conducted using the experimental software CORAL (Schaffner, 2013) with students from a large university in Australia, recruited via the ORSEE software (Greiner, 2015). After reading the instructions, subjects had to answer a couple of control questions to ensure their understanding of the game and the payoff structure.[8] Participants were then randomly assigned to the role of Proposer or Responder and kept their role during the entire experiment. At the end of the bargaining session, we also elicited participants' beliefs on the intended actions and/or beliefs of the other player.[9] At the end of the experiment, participants were randomly matched into pairs of one Proposer and one Responder, and payoffs were determined according to their choices and stated beliefs. At no time were subjects informed about the identity of their matched partner. Each session lasted approximately 45 minutes. All subjects received a fixed participation fee of \$3 and earned on average \$14.30. The full instructions can be found in Appendix A.

We used the strategy method to elicit the Responders' decisions and beliefs for each possible proposal from the Proposer. This allows us to observe how a Proposer's first proposal affects the Responder's subsequent counter-proposal.[10] To avoid experimenter demand effects, we randomly

[8]Participants who answered wrongly to some of these questions received additional explanations until the experimenter was satisfied that they understood the game.

[9]When participants made their action decisions, they did not know about the belief-estimation task, yet. Following Dufwenberg et al. (2011), we decided on this timing to avoid any influences on decisions through strategic choice making in order to subjectively simplify subsequent guesswork. This problem also partly reflects what Blanco et al. (2010) calls the risk for hedging.

[10]There are potential factors such as a reduction in incentives or a "hot" vs. "cold" effect that might affect the participants' choices by using the strategy method (Zizzo, 2010). However, the experimental evidence does not report any case in which a treat-

shuffled the presented order of first proposals.

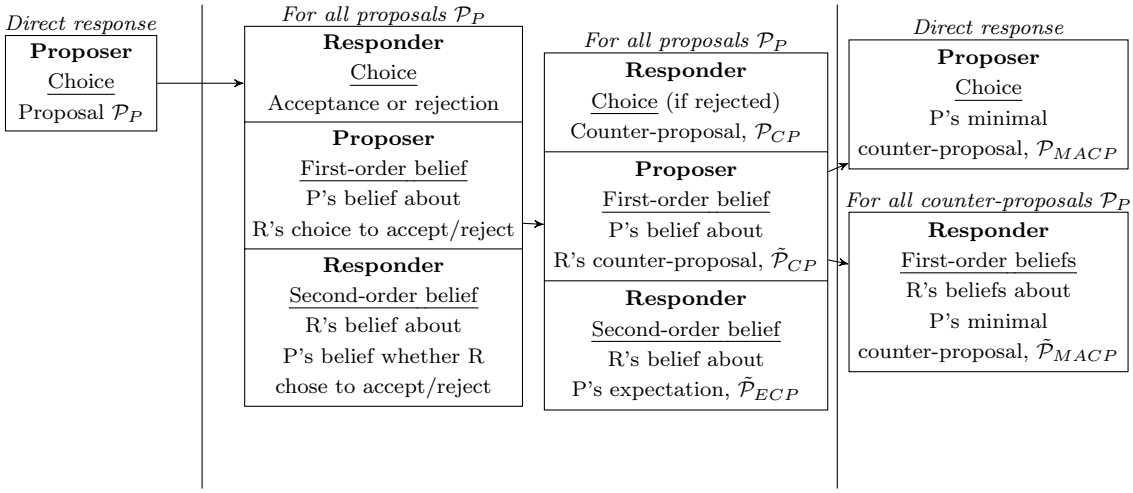


Figure 4.2: Elicitation of players' strategies and beliefs. Proposer and Responder are designated by the letters P and R, respectively.

Figure 4.2 presents all the strategies and beliefs elicited from both players. The Proposer made the decision about his initial proposition on how to split the \$10. In addition, he was asked about his minimal acceptable counter-proposal ($\mathcal{P}_{MACP}, 10 - \mathcal{P}_{MACP}$), or more specifically about his reservation value $\mathcal{P}_{MACP}$, i.e. the smallest amount which he would need to receive in order to just accept a Responder's counter-proposal. The Responder made his acceptance/rejection decision for every possible initial proposal. If he rejected a proposal, he also made a counter-proposal to suggest another split of the \$10.

In order to understand the potential dynamics initiated by a Proposer's first proposal and its effect on a Responder's decision, we elicited participants' first- and second-order beliefs about their partners' choices

ment effect is observed with the strategy method and not with the direct-response method (Brandts and Charness, 2011). Note further that although the strategy method may have an effect on the overall level of counter-proposals, the effect should be similar for all initial proposals. Our analysis will focus on the differences in reactions (rather than absolute reactions) and thus should not be affected significantly.

and beliefs. This allows us to capture if and how beliefs change as a function of a Proposer's opening proposal. Given the abstract nature of first- and second-order beliefs, we designed an innovative elicitation procedure representing the question with stylized figures of players thinking of other players.[11] Using these figures, we asked the Proposer to guess whether the Responder will accept or reject his initial proposal regarding the split of the \$10. We did so for all proposals that he could have made. In case he expected the Responder to reject the proposal, we furthermore elicited the Proposer's guess on how much money the Responder will assign him in the counter-proposal. We asked the Responder to guess how the Proposer expects him to react to each of his potential partition propositions, i.e. whether the Proposer expects him to accept or reject that particular partition. If the Responder thought that the Proposer expects a rejection, we additionally asked the Responder to guess the Proposer's belief about his counter-proposal in case of rejection ($\widetilde{\mathcal{P}}_{ECP}$). The Responder's belief about the Proposer's toughness is conveyed by his belief about the Proposer's minimal acceptable counter-proposal ($\widetilde{\mathcal{P}}_{MACP}$). This belief was elicited for all possible first proposals from the Proposer.

Participants were rewarded for correct guesses (and only for those). Given the relative complexity and large number of elicited beliefs, this procedure was chosen because it is easy to understand.[12] The Pro-

[11] First-order beliefs were elicited showing pictures of the player thinking about the action of the other player. Second-order beliefs were elicited using the same picture but now representing the other player thinking about the player's thoughts. The pictures can be found in the experimental material in Appendix A.

[12] This incentive method suffices to elicit the mode of a discrete distribution (Wilcox and Feltovich, 2000; Hurley and Shogren, 2005). Hurley and Shogren (2005) further show that this method is robust to deviations from expected utility maximization and risk neutrality. In our design, we however face the problem that participants were only asked for their belief about the other's (belief about one's) counter-proposal if they stated a belief of rejection in the first place. This may lead to a rejection guess even if subjects assign the rejection less than a 50 percent chance. Looking at our data however reveals that this potential risk is, if at all, a small problem as the acceptance

poser received \$0.50 for each correct belief about the Responder's acceptance/rejection decision. If he correctly guessed that the Responder rejected his proposal, the Proposer earned extra \$0.50 if his guess about the Responder's counter-proposal coincided with his actual counter-proposal. The Responder earned \$0.50 for each correct guess about the Proposer's minimal acceptable counter-proposal. Furthermore, the Responder received additional money if his guess about the Proposer's expectation matched the Proposer's actual expectation. The Responder earned \$0.50 for each correct guess on whether the Proposer expected him to accept the proposal. In case of a correctly expected rejection, the Responder received \$0.50 for each correct guess about the Proposer's expectation about his counter-proposal.

4.3.2 Third party observers

In order to gain richer insights in the likely beliefs and feelings of players, we elicited the views from third party observers. We recruited 59 participants to indicate their beliefs about actions and perceptions of the Proposer's intention for each of the possible proposals presented as hypothetical scenarios. By using third party observers, we avoided to make these questions about beliefs and perceptions too salient to the players during the actual experiment and avoided a potential experimenter demand effect. It also limits the risk of a false consensus effect whereby a player's beliefs can be influenced by his choices (Bellemare et al., 2011). Subjects received a flat payment of \$10 for a session of roughly 30 minutes.

The survey presented all situations a Responder could have been con-

guesses exceed the actual acceptance rate (see Figure 4.4).

fronted with to the additional participants (i.e. one scenario for each possible proposal made by the Proposer). The participants in the role of the observer were asked to consider these hypothetical situations, and to indicate their beliefs about the motive of the Proposer as well as to imagine their feelings if they were to receive these proposals as a Responder. The full questionnaire can be found in Appendix A.

4.3.3 Definitions and hypotheses

The payoff facing a Proposer at the end of the bargaining can be characterized by the Responder's final proposal, $\mathcal{P}_{FP}$:

$$\mathcal{P}_{FP} = \begin{cases} \mathcal{P}_P & \text{if the Responder accepted the first proposal,} \\ \mathcal{P}_{CP} & \text{if the Responder rejected,} \end{cases}$$

where $\mathcal{P}_{CP}$ is the counter-proposal made by the Responder after rejecting the initial proposal $\mathcal{P}_P$.

The Responder's belief about the Proposer's minimal acceptable counter-proposal is $\widetilde{\mathcal{P}}_{MACP}$. In addition, we define his belief about the Proposer's minimal acceptable proposal, $\widetilde{\mathcal{P}}_{MAP}$, as the payoff the Responder expects the Proposer to be willing to accept at the start of the game. By definition:

$$\widetilde{\mathcal{P}}_{MAP} = \min\left\{\mathcal{P}_P, \widetilde{\mathcal{P}}_{MACP}\right\}$$

And the Responder's belief about the Proposer's expected final proposal, $\widetilde{\mathcal{P}}_{EFP}$, is defined as:

$$\widetilde{\mathcal{P}}_{EFP} = \begin{cases} \mathcal{P}_P & \text{if the Responder believes the Proposer expects him to accept,} \\ \widetilde{\mathcal{P}}_{ECP} & \text{if the Responder believes the Proposer expects him to reject,} \end{cases}$$

where $\widetilde{\mathcal{P}}_{ECP}$ is the Responder's belief about the Proposer's expected counter-proposal ($\mathcal{P}_{ECP}$).

Non-cooperative game theory with own-payoff maximizing agents would predict that the first stage of the game does not give any bargaining power to the Proposer. The subgame-perfect Nash equilibrium of this game under these conditions is the same as for an ultimatum game where the Responder makes the take-it of leave-it proposal.

Hypothesis 4.1 (Common knowledge of rationality)
The Responder makes a final offer $\mathcal{P}_P = 0$.

From the results of the ultimatum game, we know that Hypothesis 4.1 is unlikely to be confirmed as players typically reject low proposals: Experimental findings show that many agents refuse very unfavourable splits due to their social preferences. We can therefore expect Proposers to refuse some unfavourable final proposals. In addition, results from dictator games show that many agents assign positive amounts to the other player even if the other cannot refuse. In our game, a selfish Responder facing a Proposer with social preferences would propose to the Proposer what he believes to be his minimal acceptable proposal. However, if the Responder has social preferences himself, he may offer the Proposer more than what he believes to be his minimal acceptable proposal.

Hypothesis 4.2 (Bargainers with social preferences)

(i) If the Responder is rational and self-regarding and he perceives that the Proposer has social preferences, then he makes a final offer $\mathcal{P}_{FP} = \widetilde{\mathcal{P}}_{MAP}$.

(ii) If the Responder has social preferences himself, then he makes a final offer $\mathcal{P}_{FP} > \widetilde{\mathcal{P}}_{MAP}$.

We are particularly interested in the role of the first proposal. Under standard assumptions, the first proposal does not carry any information and should therefore not effect the bargaining outcome.

Hypothesis 4.3 (Absence of information in first proposal) *The Responder's belief about the Proposer's minimal acceptable proposal $\widetilde{\mathcal{P}}_{MAP}$ and the Proposer's expected proposal $\widetilde{\mathcal{P}}_{EFP}$ are not affected by the Proposer's first proposal $\mathcal{P}_P$.*

Experimental evidence however suggests that cheap talk can influence players' beliefs prior to a bargaining game (Croson et al., 2003; Rankin, 2003; Tingley and Walter, 2011; Kriss et al., 2013; Anbarci et al., 2015). In contrast to Hypothesis 4.3, we therefore expect that the first proposal $\mathcal{P}_P$ may influence the Responder's beliefs and behaviour, and consider the following alternative hypothesis:

Hypothesis 4.4 (Effect of first proposals on beliefs)

(i) The Responder perceives $\mathcal{P}_P$ as conveying information about the Proposer's resolution to refuse unfavourable proposals, i.e. his minimal acceptable proposal $\mathcal{P}_{MAP}$. The more the Proposer claims in his proposal, the higher the Responder's belief about the Proposer's minimal acceptable proposal ($\widetilde{\mathcal{P}}_{MAP}$).

(ii) The Responder perceives $\mathcal{P}_P$ as conveying information about the Proposer's expectations. A higher $\mathcal{P}_P$ from the Proposer signals higher expectations regarding the final split of the pie ($\mathcal{P}_{EFP}$) and thus triggers a higher belief by the Responder ($\widetilde{\mathcal{P}}_{EFP}$).

The experimental literature also indicates that the agents' reactions depend on how they interpret the intention of the other player (e.g. Rabin, 1993; Falk and Fischbacher, 2006; Levine, 1998; Straub and Murnighan,

1995). Specifically, agents react kindly to a move perceived as triggered by a kind motive and they react unkindly to a move perceived as unkind. If first proposals have an effect on Responders' beliefs, then different first proposals may be interpreted as signalling different degrees of kindness from the Proposer. In line with Hypothesis 4.4, we assume that higher claims from the Proposer in the first proposal can be perceived as signalling a desire to get a larger part of the \$10 pie. We therefore expect low proposals $\mathcal{P}_P$ to be perceived as kinder than high proposals. Measuring kindness thereby typically requires a benchmark (Rabin, 1993). Hypothesis 4.5 proposes the Responder's belief about the Proposer's minimal acceptable final proposal ($\widetilde{\mathcal{P}}_{MAP}$) as a benchmark. This choice is motivated by the fact that a self-centred payoff-maximizing Responder would precisely offer $\widetilde{\mathcal{P}}_{MAP}$. This choice is however not critical for the rest of the analysis.

Hypothesis 4.5 (Kind behaviour)
The Responder answers positively (negatively) to the perceived kindness of the Proposer by making a final proposal ($\mathcal{P}_{FP}$) relatively higher (lower) to his belief about the Proposer's minimal acceptable final proposal ($\widetilde{\mathcal{P}}_{MAP}$).

4.4 Data and results

4.4.1 Bargaining outcomes

The outcomes of the bargaining interactions are summarized in Table 4.1. The Proposers' average first proposal $\mathcal{P}_P$ is \$5.5 or 55 percent of the surplus. In contradiction with Hypothesis 4.1, the final proposal from the Responder, $\mathcal{P}_{FP}$, is on average \$4.7. This is relatively high compared to findings from the ultimatum game (on average around of 30-40 percent of

	Mean Proposal $\mathcal{P}_P$ (SD)	Bs' Rejection Rate	Mean $\mathcal{P}_{CP}$ (SD)	Mean $\mathcal{P}_{FP}$ (SD)	As' Rejection Rate	Mean $\mathcal{P}_{MACP}$ (SD)	Mean Earnings Proposer (SD)	Mean Earnings Responder (SD)
Actually played	5.53 (1.24)	32%	4.12 (1.11)	4.73 (.79)	48%	4.68 (1.59)	4.23 (1.83)	4.25 (1.84)
For all proposals	5 (3.16)	48%	4.22 (1.40)	3.49 (1.88)	n/a	n/a	n/a	n/a

Table 4.1: Summary statistics.

the pie, see Camerer, 2003a). Payoffs are on average balanced between the Proposers and the Responders with \$4.23 and \$4.25, respectively. Responders' relatively high average earnings are driven by two mechanisms: First, even though first proposals are costless to reject, they are accepted most of the time, this is the case in 68 percent of the bargaining situations. Second, even when first proposals are rejected, the counter-proposal still tends to be substantial. Rejected first proposals $\mathcal{P}_P$ are on average \$6.64 and the subsequent counter-proposals $\mathcal{P}_{CP}$ are on average \$4.1. Even though the first proposal should not give bargaining power to the Proposer under standard assumptions, we observe that being able to make a first proposal allows the Proposer to get a share of the pie which is higher than in the ultimatum game where the player receiving the last offer also has no bargaining power.

4.4.2 Proposers' choices

As can be seen in Figure 4.3, the large majority of Proposers makes a 50:50 split-proposal. Nobody requests less than \$4 and only less than 13 percent make proposals different to \$5 or \$6.

Although the final level of counter-proposals is on average above \$4, counter-proposals are rejected by the Proposer in 52 percent of the time.

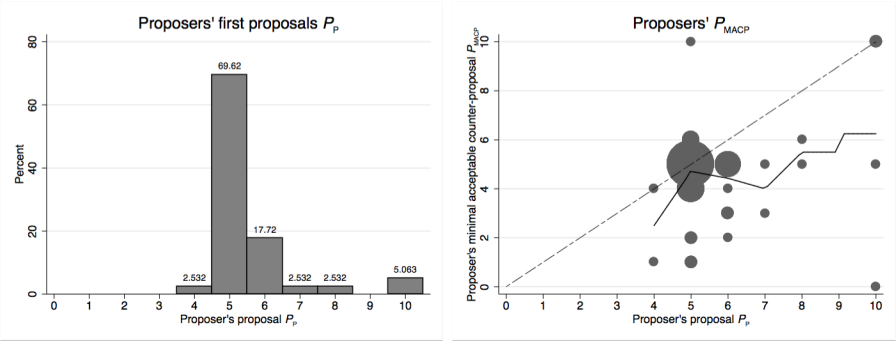


Figure 4.3: The Proposers' decisions. The left panel shows the distribution of the first proposals. The right panel shows Proposers' minimal acceptable counter-proposals, $\mathcal{P}_{MACP}$, depending on their initial proposal.

When the counter-proposal is below \$5, it is rejected in over 80 percent of times which is high compared to ultimatum games (Camerer, 2003a, lists 16 studies with average rejection rates typically in the range of 10-30 percent). The high level of rejections is driven by the Proposers' high average level of minimal acceptable counter-proposals ($\mathcal{P}_{MACP}$) of \$4.7. The average $\mathcal{P}_{MACP}$ is rather large compared to the average level of minimal acceptable proposals in typical ultimatum studies, which is around 20-30 percent of the pie size (e.g. Blount, 1995; Straub and Murnighan, 1995).[13]

While the $\mathcal{P}_{MACP}$ is rather large, the first proposal is not a simple reflection of it. All but two participants choose to make a first proposal equal to or above \$5. For most Proposers who make a proposal $\mathcal{P}_P = 5$, we observe $\mathcal{P}_{MACP} = \mathcal{P}_P = 5$. A request of a fair split is therefore a rather informative signal that the Proposer is likely to reject an unequal split. For Proposers who make a proposal $\mathcal{P}_P > 5$, we observe that almost all the time $\mathcal{P}_P > \mathcal{P}_{MACP}$. Nevertheless, there is a positive relationship

[13]It could be that the possibility of making an initial proposal evokes an increased sentiment of entitlement and/or reciprocity-sensitivity. Moreover, it could be driven by the Proposers' desire to be consistent with their stated proposal as suggested by the findings of Cialdini et al. (1995).

between the Proposers' first proposals $\mathcal{P}_P$ and their minimal acceptable counter-proposals $\mathcal{P}_{MACP}$. This pattern indicates that demanding proposals $\mathcal{P}_P > 5$ are not a strong indication of the Proposer's toughness of his bargaining stance although the level of the proposal contains some information about the likelihood that the Proposer rejects low counter-proposals.

Overall, this pattern supports part (i) of Hypothesis 4.4 whereby the first proposal conveys, in practice, some information about the Proposer's $\mathcal{P}_{MACP}$. It is also compatible with the literature on bargaining with reputation where a bargainer can send a signal of commitment to reject future proposals below some threshold (Abreu and Gul, 2000; Wolitzky, 2012; Embrey et al., 2014). In our sample, a substantial proportion of proposers adopt an "obstinate" strategy where they commit to refuse counter-proposal below \$5 while making a proposal of \$5. In such a situation, the existence of such obstinate players gives weight to initial signals and allows players sending the signal to claim a substantial part of the pie. We observe indeed that the final payoffs of Proposers and Responders are almost equal in our design which is in stark contrast to findings from ultimatum games even though our Proposer has under standard assumptions the same bargaining power as the recipient of the ultimatum offer. The possibility to make a first proposal coupled with the existence of a substantial amount of obstinate players improves the bargaining power of the Proposer significantly.

4.4.3 Responders' reaction to the opening proposal

Most Responders accept any proposal where the Proposer requests \$5 or less. For initial proposals $\mathcal{P}_P$ above \$5, the acceptance rate falls sharply

and quickly reaches nearly zero. The acceptance rate of the Responder (as well as the associated beliefs from the two players) are represented in Figure 4.4. The Proposers' expectations regarding the Responders' acceptance decision as well as the Responders' belief about the Proposers' expectations are on average quite accurate. Players do not behave as if they are payoff-maximizing agents with common knowledge of rationality. Instead, the equal payoff allocations appears as a benchmark with proposals from the Proposer requesting more than 50-50 being much more likely to be rejected.

Result 4.1 (Acceptance of initial proposals)

Most Responders accept non-zero proposals $\mathcal{P}_P$ which are below or equal to \$5. Their acceptance rate falls quickly for proposals requesting more than the equal payoff allocation. For $\mathcal{P}_P$ of \$7 and above, it is close to zero.

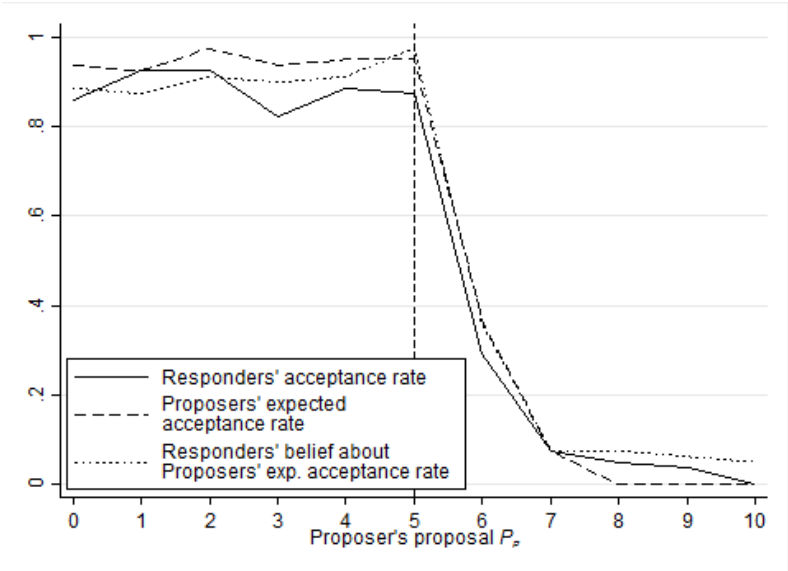


Figure 4.4: a) The Responders' acceptance rate; b) The Proposers' average expected acceptance rate; c) The Responders' average belief about the Proposers' expected acceptance rate.

Closely related to the Responder's acceptance decision is his final proposal ($\mathcal{P}_{FP}$). As can be seen in Figure 4.5, the Responders' average $\mathcal{P}_{FP}$ follows closely the 45 degree line up to a proposal of approximately \$5 (reflecting the high acceptance rate for low proposals). Note that it lies slightly above it up to a proposal of \$4 and from then onwards slightly below. Note further that it never reaches the equal payoff allocation of \$5. The observed severe kink at a proposal of \$5 reflects the sudden drop in the acceptance rate.

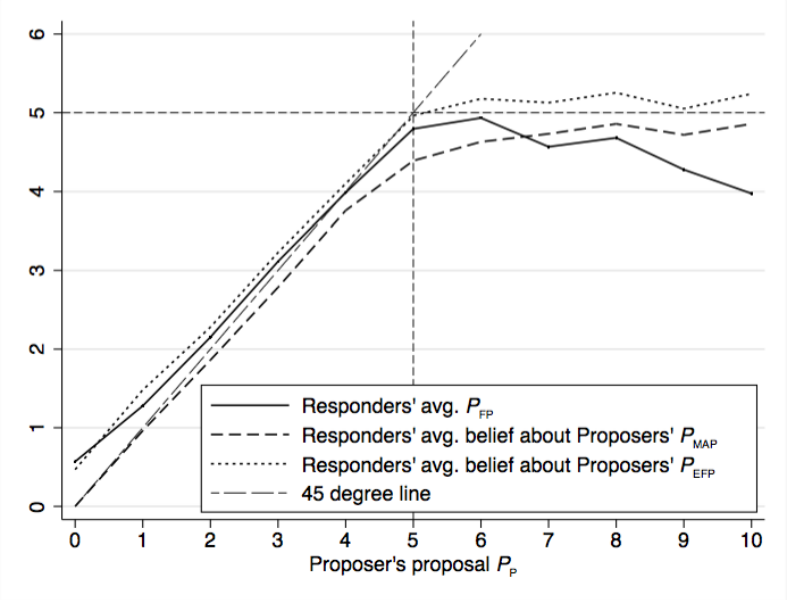


Figure 4.5: Responders' a) average final proposal, $\mathcal{P}_{FP}$, alongside their average beliefs about b) the expectation of the Proposer, $\widetilde{\mathcal{P}}_{EFP}$, and about c) the Proposer minimal acceptable proposal, $\widetilde{\mathcal{P}}_{MAP}$, depending on the Proposer's initial proposal.

Most noticeable is the kink in the curve with the average final proposal declining as the first proposal reaches more than 60 percent of the amount to split. This downward slope is significant using an OLS regression for proposals above \$5. Table 4.2 presents the result of this regression.

	$\mathcal{P}_{FP}$ (Robust SD)	$\widetilde{\mathcal{P}}_{MAP}$ (Robust SD)	$\widetilde{\mathcal{P}}_{EFP}$ (Robust SD)
Constant	6.261***	4.408***	5.129***
	(.286)	(.249)	(.326)
Proposal $\mathcal{P}_P$ (> 5)	-.222***	.044	.005
	(.041)	(.035)	(.050)

Estimated on the subsample of proposals $\mathcal{P}_P > 5$, N=395.
Standard errors in brackets. $^*p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$.

Table 4.2: Slope estimations for the Responder's $\mathcal{P}_{FP}$, $\widetilde{\mathcal{P}}_{MAP}$ and $\widetilde{\mathcal{P}}_{EFP}$ as a function of the Proposer's first proposal (for $\mathcal{P}_P >5$).

Result 4.2 (Final proposals as a function of first proposals)
The Responders' average final proposal increases with the opening proposal if this proposal is below 50 percent of the pie. It decreases with the proposal if it is above 60 percent of the pie.

We now turn to the question how the Proposer's first proposal $\mathcal{P}_P$ impacts the Responders' $\widetilde{\mathcal{P}}_{MAP}$ and their $\widetilde{\mathcal{P}}_{EFP}$. The associated data is displayed in Figure 4.5. When the Proposer's first proposal is below or equal to five, most Responders accept such proposals and also believe that they are expected to do so. Hence, $\widetilde{\mathcal{P}}_{EFP} \approx \mathcal{P}_P$. However, once proposals exceed the equal payoff split, i.e. when $\mathcal{P}_P > 5$, Responders' $\widetilde{\mathcal{P}}_{EFP}$ becomes a function of the Responder's belief about what the Proposer expects as a counter-proposal. $\widetilde{\mathcal{P}}_{MAP}$ follows a similar course: $\widetilde{\mathcal{P}}_{MAP} \approx \mathcal{P}_P$ for $\mathcal{P}_P < 5$ but the Responders' average belief about the minimal counter-proposal the Proposer would accept, $\widetilde{\mathcal{P}}_{MAP}$, increasingly deviates from $\mathcal{P}_P$ for larger opening proposals.

Interestingly, proposals above \$5 do not seem to impact the Responders' beliefs on average. They are interpreted by the Responder as indicating a $\mathcal{P}_{MAP}$ and a $\mathcal{P}_{EFP}$ of around five, no more. In that way, proposals requesting more than the equal split seem to have the same

effect as cheap talk.[14] We verify the visually obtained result by estimating the effect of first proposals above five on the Responders' $\widetilde{\mathcal{P}}_{MAP}$ and $\widetilde{\mathcal{P}}_{EFP}$. We find that the Responders' beliefs do not change significantly as a function of $\mathcal{P}_P$ for $\mathcal{P}_P > 5$ (see Table 4.2).

Result 4.3 (Demanding proposals and Responders' average beliefs) *Proposals above the equal payoff split are <u>not</u> interpreted as:*

(i) Signalling a credible indication of a tough bargaining stance (high $\widetilde{\mathcal{P}}_{MAP}$).

(ii) Signalling a higher payoff expectation from the Proposer (high $\widetilde{\mathcal{P}}_{EFP}$).

Notably, the decrease in the final proposal $\mathcal{P}_{FP}$ for initial proposals above \$5 does not seem to be driven by these beliefs as they stay roughly constant.[15] However, hiding behind the absence of effect of first proposals on average beliefs, we observe a change in the distribution of Responders' beliefs. Figure 4.6 displays these distributions. For first proposals above \$5, Responders' beliefs become more dispersed.

A closer look at the relationship between Responders' average final proposals and their average beliefs shows that final proposals exceed the Responders' belief about the Proposer's $\mathcal{P}_{MAP}$ up to an initial proposal of approximately \$7 (see Figure 4.5). This result could suggest a small impact of distributional preferences or positive reciprocity. In several cases, the Responder accepts a reasonable split-proposition even though he could try to claim more for himself. Very modest (kind) opening proposals from the Proposer (up to approximately \$4) are sometimes even

[14] Note that high initial proposals also do not signal a "gamesman" with a very low $\mathcal{P}_{MAP}$.

[15] The difference in slopes between the Responders' average $\mathcal{P}_{FP}$ and the Responders' average $\widetilde{\mathcal{P}}_{MAP}$ is significant ($p = 0.006$).

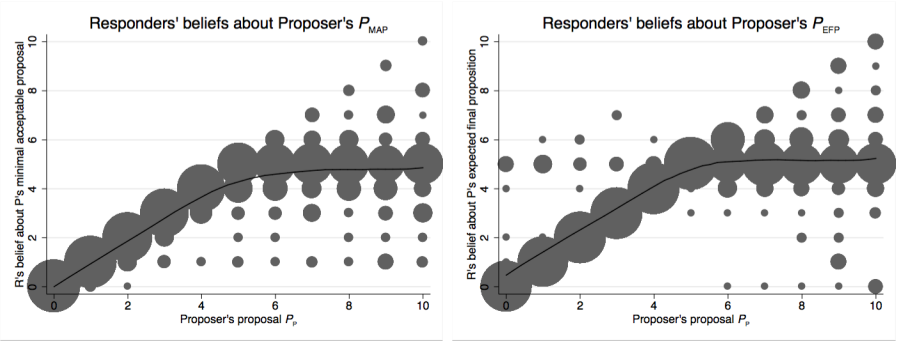


Figure 4.6: Responders' beliefs about the Proposer's minimal acceptable proposal, $\widetilde{\mathcal{P}}_{MAP}$ (left panel), and about his expected final proposal, $\widetilde{\mathcal{P}}_{EFP}$ (right panel).

followed by a counter-proposal offering the Proposer more than what he had asked for ($\mathcal{P}_{FP}$ above the 45 degree line). Once first proposals go beyond \$7, though, the Responders' $\mathcal{P}_{FP}$ falls on average short of the $\widetilde{\mathcal{P}}_{MAP}$. The Responders respond to high proposals by making counter-proposals which are on average lower than what they believe the Proposer's $\mathcal{P}_{MAP}$ to be. By doing so, they seem to punish the Proposer at their own cost as according to their own beliefs their final proposal is likely to lead to a rejection from the Proposer and therefore to a null payoff for both players.[16] This pattern gets stronger as proposals tend towards the full amount of the pie. Figure 4.7 displays the pattern of such punishing behaviour. The left panel shows the difference between $\mathcal{P}_{CP}$ and $\widetilde{\mathcal{P}}_{MAP}$ for each initial proposal. The right panel shows the proportion of punishing counter-proposal ($\mathcal{P}_{CP} < \widetilde{\mathcal{P}}_{MAP}$) for each initial proposal. For a very high initial $\mathcal{P}_P$, more than 40 percent of the counter-proposals can be considered as punishing.

[16]This interpretation is supported by Schweinsberg et al. (2012)'s psychological study which found evidence that extreme requests in negotiation may lead to the other party walking away.

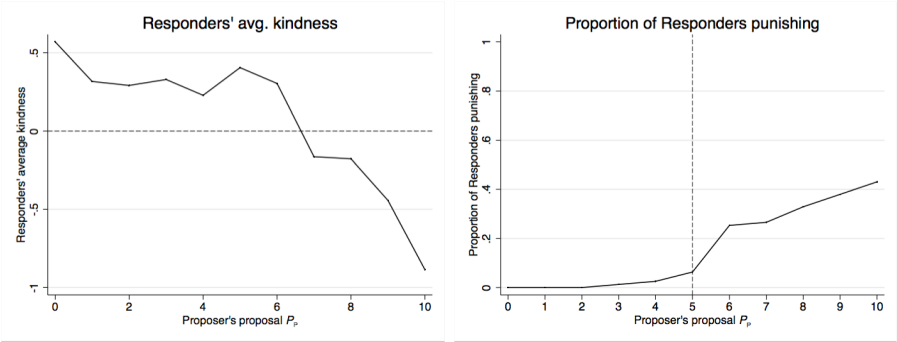


Figure 4.7: Responders' punishing behaviour. The left panel shows the difference between the Responders' average final proposal and their average belief about the Proposer's $\mathcal{P}_{MAP}$. The right panel shows the proportion of Responders making a punishing counter-proposal for each proposal. The Responder is said to punish the Proposer if $\mathcal{P}_{CP} < \mathcal{P}_{MAP}$.

Result 4.4 (Punishing demanding proposals)

The Responders' $\mathcal{P}_{FP}$ exceeds on average their $\widetilde{\mathcal{P}}_{MAP}$ for low first proposals from the Proposer. For higher first proposals, the Responders' average $\mathcal{P}_{FP}$ decreases and drops below their average $\widetilde{\mathcal{P}}_{MAP}$. Responders seem willing to terminate the bargaining without an agreement as a consequence of the Proposer's first proposal.

4.4.4 External observers' perceptions

Observers' expectations about players' choices and beliefs in the game resemble the actual pattern of answers observed in the experiment (see Figure 4.8). This result suggests that they have a good understanding of the game and the likely state of mind of the Proposer and Responder in each situation.

We also asked the observers about their perceived intentions and perceptions of the different bargaining situations. Figure 4.9 shows what observers consider to be the likely motive driving each possible level of a Proposer's opening proposal. Unsurprisingly, most observers regard the

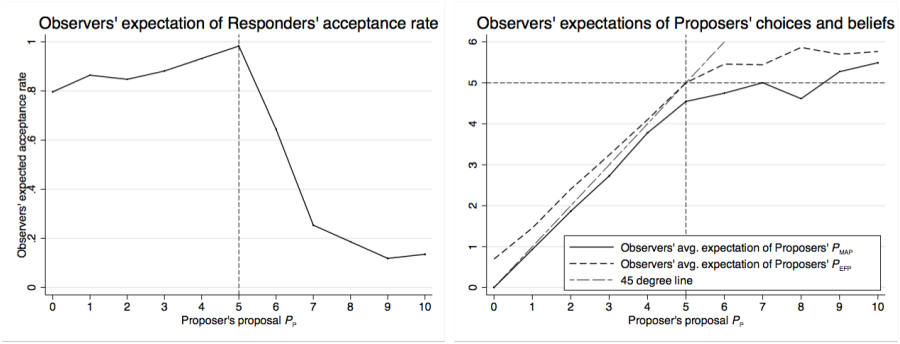


Figure 4.8: Observers' expectations about the Responders' acceptance rate (left panel) and about Proposers' expected final proposal, $\mathcal{P}_{EFP}$, and their minimal acceptable proposal, $\mathcal{P}_{MAP}$, (right panel).

Proposer as fair when making a proposal $\mathcal{P}_P = 5$. Proposals above \$6 are perceived as indicating selfishness. However, a proposal of $\mathcal{P}_P = 6$ is not yet perceived as selfish by many observers. Similarly, half of the observers interpret offers $\mathcal{P}_P \geq 6$ as indicating a desire to secure more than half of the \$10. Proposers making proposals with a very large share for themselves, $\mathcal{P}_P \geq 9$, are even classified as nasty by a third of the observers. Observers generally do not tend to interpret first proposals as being used strategically to suggest to the Responder that the Proposer has a higher $\mathcal{P}_{MAP}$ than it is. The proposals perceived as the most "reasonable" are $\mathcal{P}_P = 5, 6, 7$, with lower and higher requests being harder for the observers to make sense of.

Additionally, we asked observers about their feelings if they were faced with each possible opening proposal. Figure 4.10 represents the fraction of observers experiencing a particular feeling. Observers tend to feel good and happy for proposals $\mathcal{P}_P \leq 5$ and angry and insulted for proposals $\mathcal{P}_P \geq 7$. Interestingly, the proposal evoking a neutral feeling by most observers is $\mathcal{P}_P = 6$. This can be understood as a level for which they feel neither happy as it is beyond the equal payoff nor insulted/angry yet

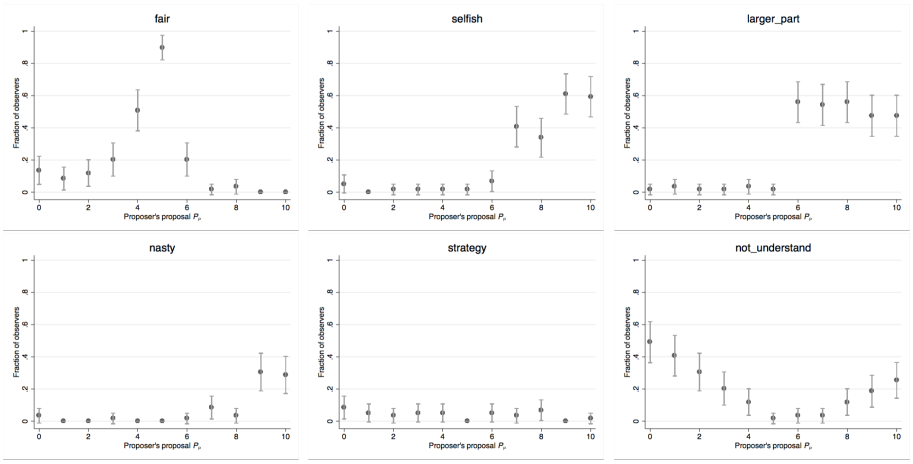


Figure 4.9: Observers' views on the motives driving the Proposer's different opening proposals.

as it is not a large deviation from it.

4.4.5 What drives these intention-based preferences?

Our experimental design allows us to investigate the mechanism behind the Responder's reaction to the Proposer's proposal.

According to the reciprocity approaches by Rabin (1993) as well as Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006), perceived kindness in zero-sum games depends crucially on the Responder's second-order belief: The more the Responder believes the Proposer expects to accrue for himself from the final payoff allocation, the less kind the Proposer's choice is perceived. Hence, given a certain choice, the higher the Responder's belief about the Proposer's payoff expectation, the less kind the Proposer's choice is interpreted and in turn the less kind the Responder's response should be. Our elicitation of first- and second-order beliefs in this experiment allows us to investigate the empirical predictions of psychological game theory in this bargaining contest.

We observed that first proposals have an effect on the Responder's

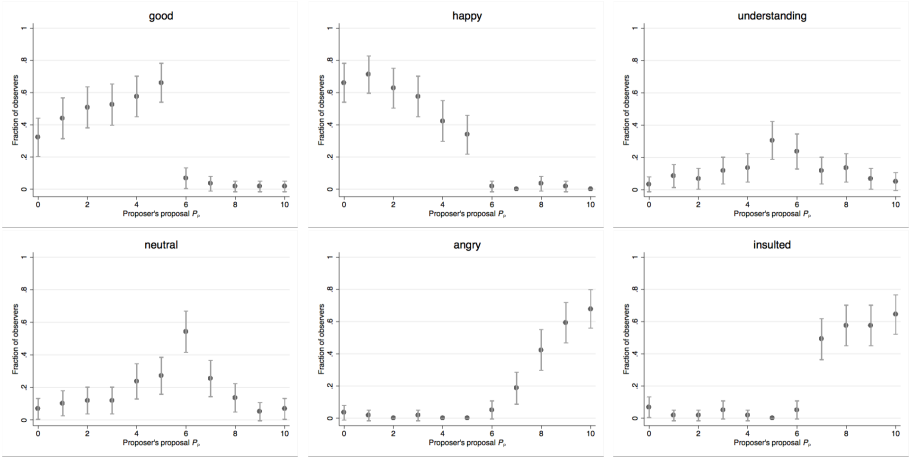


Figure 4.10: Observers' feelings when faced with the Proposer's different opening proposals.

belief about the Proposer's expectations. While, on average, proposals above \$5 are associated with expectations of around \$5, Figure 4.6 revealed that Responders' beliefs are quite dispersed. We therefore test whether punishing counter-offers are Responders' answer to their belief that the Proposer expects to get more than the equal payoff split resulting from his proposal. In Table 4.3 (column 1), we regress the Responder's final proposal on the Proposer's opening proposal and his beliefs – $\widetilde{\mathcal{P}}_{MAP}$ and $\widetilde{\mathcal{P}}_{EFP}$. We find a positive but insignificant effect of $\widetilde{\mathcal{P}}_{MAP}$. We cannot rule out that the Responder does not react to his belief $\mathcal{P}_{MAP}$ when making a counter-proposal. The effect of $\widetilde{\mathcal{P}}_{EFP}$ on Responders' $\mathcal{P}_{FP}$ is surprisingly positive (not negative). This result is in contradiction with the idea that reciprocity concerns may be driving the drop in counter-proposals for high first proposals from the Proposer: Since Responders' $\widetilde{\mathcal{P}}_{EFP}$ is on average > 5 for proposals $\mathcal{P}_P > 5$ (Figure 4.6), models of reciprocity would predict that negative reciprocity should prevail for $\mathcal{P}_P > 5$ and Responders' counter-proposals should be lower, not higher, the higher they believe the Proposer's expectation to

	(1) $\mathcal{P}_{FP}$	(2) Punishing $\mathcal{P}_{CP}$
Constant	4.929***	-0.664***
	(0.597)	(0.124)
$\mathcal{P}_P$	-0.229***	0.040***
	(0.043)	(0.011)
$\widetilde{\mathcal{P}}_{MAP}$	0.143	0.168***
	(0.089)	(0.015)
$\widetilde{\mathcal{P}}_{EFP}$	0.137*	-0.023
	(0.061)	(0.016)

For proposals $\mathcal{P}_P > 5$, N=395.
Robust standard errors in brackets.
$^*p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$.

Table 4.3: Slope estimations for (1) Responder's level of the $\mathcal{P}_{FP}$, (2) Responder's decision to punish (binary).

be. Even more surprising, we find that, for high first proposals, the level of the proposal influences the Responder's kindness negatively even after controlling for $\widetilde{\mathcal{P}_{PMAP}}$ and $\widetilde{\mathcal{P}}_{EFP}$. This pattern suggests that the level of the first proposal has an effect which is not reducible to its impact on the Responder's belief about the Proposer's expectation (and/or his $\widetilde{\mathcal{P}}_{MAP}$).

In Table 4.3 (column 2), we take a closer look at punishing counter-proposals. We regress the Responder's decision to make a punishing counter-proposal on the Proposer's opening proposal and his beliefs – $\widetilde{\mathcal{P}}_{MAP}$ and $\widetilde{\mathcal{P}}_{EFP}$. The most important insight to be gained is the insignificance of the Responder's $\widetilde{\mathcal{P}}_{EFP}$: The Responder's decision whether to punish or not does not seem to be driven by negative reciprocity (as considered by psychological game-theoretic models) as it is independent of the Responder's belief about the Proposer's expected final proposal.

Result 4.5 (Final proposals not driven by beliefs on Proposer's expectation)

(i) The level of the final proposal following a first proposal above \$5 is not negatively correlated with $\widetilde{\mathcal{P}}_{EFP}$.

(ii) The probability to make a punishing counter-proposal, $\mathcal{P}_{CP} < \widetilde{\mathcal{P}}_{MAP}$, *is not correlated with* $\widetilde{\mathcal{P}}_{EFP}$.

Notably, the decreasing slope in counter-proposals for high first proposals also cannot be explained by distributional preferences which would predict counter-proposals staying close to a constant (typically \$5) reflecting the preferred split (typically 50-50).

Turning to the data elicited from the third party observers allows us to gain more insights into the drivers behind Responders' reactions. The observers were asked more emotionally weighted questions about their perception of the game. In particular, they were asked about their perceived type or intention of the Proposer and his actions. They were also asked how they would feel if they were in the shoes of the Responder and whether they would want to make a punishing counter-offer or even stop the game without any counter-offer (leading to a payoff of zero for both players).

Table 4.4 presents the results of regressions where observers' beliefs about $\mathcal{P}_{MAP}$, $\widetilde{\mathcal{P}}^{O}_{MAP}$, and about $\mathcal{P}_{EFP}$, $\widetilde{\mathcal{P}}^{O}_{EFP}$, as well as perceived types/ intentions and associated feelings are used to explain answers to the question about the experienced feelings of happiness, anger and insult, and the decision to end the game or punish the Proposer. We find that $\widetilde{\mathcal{P}}^{O}_{MAP}$ and $\widetilde{\mathcal{P}}^{O}_{EFP}$ both influence the feeling of happiness (1-2). But only a high $\widetilde{\mathcal{P}}^{O}_{MAP}$ is associated with anger (3-4). When controlling for beliefs about the other's type, neither $\widetilde{\mathcal{P}}^{O}_{MAP}$ nor $\widetilde{\mathcal{P}}^{O}_{EFP}$ predicts the feeling of being insulted (6). Looking at the choice to end the game or punish the

Proposer, we observe that $\widetilde{\mathcal{P}}^{O}_{EFP}$ is most often not significant while $\widetilde{\mathcal{P}}^{O}_{MAP}$ is significant in most cases (7-12).

Additionally, we looked at other perceived intentions/types as factors evoking certain feelings and negative actions (ending the game or making punishing counter-offers). We find that the feeling of being insulted is significant both in predicting punishing counter-proposals and decisions to end the game. The belief that the Proposer wants to secure the largest part of the \$10 is in turn the most predictive of the feeling of being insulted. The feeling of kindness is significant when predicting punishing counter-proposals, but not when predicting decisions to end the game.

Taken together, these results suggest that preferences over intentions are driving the negative reaction to high proposals from the Proposer. We observe strong negative emotional reactions for demanding proposals $\mathcal{P}_P > 6$ and a substantial propensity to punish the Proposer for such proposals. The desire to punish is linked with the observers' perception of the Proposer's intention. The absence of a clear link between $\widetilde{\mathcal{P}}^{O}_{EFP}$ and negative emotions as well as punishing behaviour does not point to an explanation relying on the positive/negative reciprocity models from psychological game theory. Alternative explications of intention-based preferences such as models where agents are spiteful against non-altruistic agents (Levine, 1998) or have a "wounded pride" when receiving low proposals (Straub and Murnighan, 1995; Yamagishi et al., 2012) seem better able to account for the observed behavioural pattern.

Table 4.4: Observers' views. Models (1)-(6) regress the observer's feelings on his beliefs. Models (7)-(12) regress the observer's desire to end the game or punish the Proposer on his beliefs and feelings.

	(1) Happy	(2) Happy	(3) Angry	(4) Angry	(5) Insulted	(6) Insulted	(7) End	(8) End	(9) End	(10) Punish	(11) Punish	(12) Punish
$\widetilde{\mathcal{P}}_{MAP}$	-0.079***	-0.048***	0.057***	0.038***	0.052***	0.022*	0.015	0.007	0.001	0.065***	0.050***	0.033*
	(0.011)	(0.009)	(0.010)	(0.011)	(0.009)	(0.009)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
$\widetilde{\mathcal{P}}_{EFP}$	-0.042***	-0.031***	0.005	0.000	0.019	0.014	-0.003	-0.004	-0.008	0.020	0.015	0.009
	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)	(0.010)	(0.013)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)
Larger Part		-0.143***		0.143**		0.313***		0.031			0.068	
		(0.030)		(0.052)		(0.052)		(0.054)			(0.052)	
Kind		0.317***		-0.146***		-0.143***		-0.053			-0.150**	
		(0.061)		(0.029)		(0.026)		(0.055)			(0.044)	
Look tough		0.198*		-0.109		-0.124		-0.278***			-0.092	
		(0.085)		(0.058)		(0.064)		(0.037)			(0.081)	
Insulted									0.210**			0.137
									(0.076)			(0.070)
Happy									-0.010			-0.166***
									(0.064)			(0.041)
Angry									0.049			0.199**
									(0.077)			(0.066)
Constant	0.774***	0.546***	-0.044	0.060	-0.038	0.062	0.233**	0.285***	0.252**	-0.016	0.093*	0.128*
	(0.052)	(0.064)	(0.041)	(0.050)	(0.041)	(0.043)	(0.073)	(0.073)	(0.082)	(0.037)	(0.043)	(0.059)
Observations	649	649	649	649	649	649	649	649	649	647	647	647

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.4.6 What is a the profit-maximizing opening proposal?

We have seen that Responders will accept low proposals $\mathcal{P}_P$ but punish proposals which are too high above the equal split. In that context, an interesting side question is what is the "best" or rather most profitable opening proposal. Figure 4.5 revealed that the highest final proposal is on average made following a proposal slightly above the equal payoff split. Since averages mask individual differences, we additionally look at the choice distributions. Figure 4.11 displays the boxplots of Responders' final proposals for each first proposal from the Proposer. For proposals up to a value of five, the large majority of Responders accept the Proposer's proposal. Hence the boxplot contracts to a line at the exact value of the proposal. The most interesting observations can be made for proposals between five and seven. While the mode of Responders' final proposals remains at five, the fairly homogeneous behaviour of Responders trails off. First proposals of six are still accepted by a decent number of players but are mostly followed by a counter-proposal of five so that the 75^{th} percentile moves upwards. For proposals of seven and larger, this trend is reversed. Indeed, the 75^{th} percentile returns to a level of five and the 25^{th} percentile declines to four: A substantial number of Responders rejected and chose a counter-proposal below five.

On aggregate, these responses to a proposal of six deliver an average payoff for the Proposer which is higher than a proposal of five. The optimal initial payoff proposal from the Proposer's point of view is therefore not the equitable 50:50 split but it lies slightly above it. Yet, this is a slightly more risky choice which might explain why the large majority of players in the role of the Proposer chose a proposal of five and only

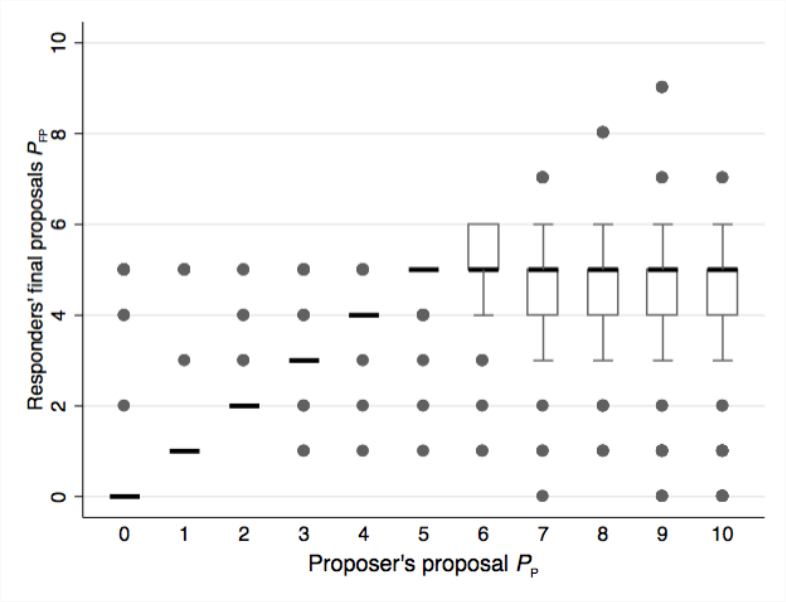


Figure 4.11: Boxplot of Responders' final proposals for each first proposal from the Proposer.

roughly 20 percent went for the on average payoff-maximizing proposal of six (see Figure 4.3). Note that while six is the optimal first proposal, it is associated with an expected final payoff of almost exactly five (see Figure 4.5). Therefore, asking for an amount slightly above the equal payoff (fairness) split was the best way to ensure such an equal split in the end.

Result 4.6 (The optimal first proposal is just above the equal payoff split)

A Proposer's maximal expected final proposal can be obtained by making a first proposal of six which leads on average to a slightly better final proposal than a first proposal of five.

The data from the third party observers give additional valuable insights into why this is the case. For each proposal, the observers were asked how they would feel if faced with such a proposal. While a proposal of six did not make the observer happy, it also did not make him angry.

In fact, it is for proposals of six that the observers think they would feel most neutral as a Responder.

4.5 Summary and discussion

We have investigated the effect of opening proposals on the outcome of a negotiation. We designed a double-round alternating-offer bargaining game mimicking the typical start of a negotiation process, with one bargainer making a proposal for a split of a $10, which the other bargainer can accept or reject to make a counter-proposal. This design allows us to observe how a bargainer receiving the opening proposal in a negotiation reacts to it. While the Proposer's initial proposal should not influence the final outcome under classic assumptions, we find that it nevertheless influences the Responder's behaviour with the Responders' average final proposal following an inverse U-shape: The Responders' final proposal is maximal for first proposals close to the even payoff split and smaller both for lower and for higher proposals from the Proposer.

In line with the literature on bargaining with reputation (Abreu and Gul, 2000; Wolitzky, 2012; Embrey et al., 2014), we interpret this pattern as resulting from the informational effect of the first proposals. We observe that most Proposers ask for the equal payoff and that it is a credible signal of their willingness not to accept any lower amounts as they set their minimal acceptable proposal also at the equal payoff. Responders rightly react to Proposer's first proposal and in most cases do not try to make a lower counter-proposal when the first proposal is equal or below the equal payoff split. However, for proposals requesting more than the equal split, Proposers are mostly willing to accept lower counter-proposals. Responders accurately perceive that proposals above

the equal split do not signal their true minimum acceptable amount. Though, such proposals are not well received by Responders who tend to make punishing counter-proposals. A survey from external observers shows that proposals requesting too much are likely to trigger Responders' anger and unhappiness at the perception that the Proposer is trying to secure more than the equal split.

These results indicate that opening proposals can send two different signals. Proposals equal or below the equal split are informative about the Proposers willingness to refuse any lower amount. They may therefore signal that the Proposer is from an obstinate type and induce the Responder to accept the proposal. Proposals requesting more than the equal split are not credible signals of stubbornness. However, they signal that the Proposer would be willing to secure more than the equal split and that he is from an unkind type in terms of social preferences. This induces the Responder to react with spitefulness to a first proposal requesting more than the equal split. The inverse U-curve is the result of these two effects: Proposals up to the equal split signal an obstinate type who will likely reject anything below the requested amount and proposals above the equal split signal an unkind type which triggers a spiteful response from the Responder.

Our findings nicely unite two conflicting views in the existing psychological literature on first proposals in negotiations. On the one side, it is argued that a high opening proposal results in a higher individual outcome because it works as an anchor. Tversky and Kahneman (1974) have shown that an initial salient piece of information is used to make subsequent judgements. And because behavioural adjustments tend to be insufficient, irrespective of the level of the anchor, high opening proposals bias the final outcome in the direction of the proposal (Galinsky and

Mussweiler, 2001). The second view suggests that bargainers initiating the negotiation with a reasonable proposal achieve the more favourable outcomes because extremely high opening proposals may sour the atmosphere and endanger the agreement (Raiffa, 1982; Schweinsberg et al., 2012). We find support for both hypotheses: The highest bargaining outcome is not obtained for a first proposal requesting the equal split, but for a slightly higher request which increases the chances to get the equal split or above. But proposals with too large requests trigger a negative reaction from the Responder's preference over intentions.

A large body of economic research has shown that bargainers' distributional preferences play a substantial role in their decision to accept a deal or not. This study adds to this evidence by showing that another type of social preferences, intention-based preferences, also plays a role. As a consequence, the success or failure of a negotiation does not only depend on the final proposal on the table but also on the emerging dynamics of the bargaining process. The intermediary proposals made during a negotiation can be interpreted by the other bargainer as suggesting either kind and compromising intentions or unkind and tough ones. And the perception of such intentions can, in turn, influence the final outcome of the bargaining process. For this reason, as suggested by the quote of Raiffa in a classical book on negotiation at the beginning of this article, it is not the best strategy to always be as tough as possible in a negotiation.[17]

The role played by intention-based preferences in bargaining suggests that striking a good bargain is a balancing act requiring not to be too soft

[17]The importance of (reciprocal) fairness in bargaining processes echoes the role it has been found to play in other economic activities such as firms' price setting decisions (Kahneman et al., 1986), wage negotiations (Kahneman et al., 1986; Campbell III and Kamlani, 1997), the contribution to public goods (Sugden, 1984) or contract enforcement (Fehr et al., 1997).

(as it is not often rewarded) and not too tough (as it is often punished). The field of Negotiation, taught in Business Schools, investigates the role of soft skills in negotiations: It asks how the interaction process during the negotiation can be used to enhance the likelihood to reach a successful agreement. The present research suggests that economists can meaningfully venture into this aspect of economic behaviour using inter alia the insights of models of intention-based preferences as well as signalling game theory.

Chapter 5

Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games

Abstract[1]

We investigate the role of intentions in two-player two-stage games. For this purpose, we systematically vary the set of opportunity sets the first mover can chose from and study how the second mover reacts not only to opportunities of gains but also of losses created by the choice of the first mover. We find that the possibility of gains for the second mover (generosity) and the risk of losses for the first mover (vulnerability) are important drivers for second mover behaviour. Efficiency concerns and an aversion against violating trust, on the other hand, seem to be far less important motivations. We also find that second movers compare the actual choice of the first mover and the alternative choices that would have been available to him to allocations that consist of equal material payoffs.

[1]This is a joint study with Rudolf Kerschbamer and Lionel Page.

5.1 Introduction

Other-regarding preferences capture people's valuation not only for their own material resources but also for the material payoffs of other individuals as well as the perceived kindness of others' behaviour. The theoretical literature on such preferences can be divided into two broad classes: models with distributional (unconditional) other-regarding preferences and models with intention-based (conditional) other-regarding preferences.

The distributional preference approach focuses on preferences over allocations of resources which are driven by distributional properties of the allocations. The altruism models by Andreoni and Miller (2002) and by Cox et al. (2007) fall into this category, as well as the models of inequality-aversion by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), and the model of altruism and spite by Levine (1998).[2]

The intention-based approach, on the other hand, tries to explain findings neither consistent with self-regarding preferences nor in line with existing models of distributional concerns by agents' desire to react to others' intentions. In this strand, a second mover's preferences in a two-person two-stage game typically become more or less benevolent depending on the perceived "kindness" of the first mover, and kindness is interpreted as generosity.

Two approaches have been proposed to investigate intention-based preferences theoretically. First, in psychological game theory, a player evaluates another person's kindness by forming beliefs on what the other person believes the consequences of his choice to be (see Rabin, 1993; Dufwenberg and Kirchsteiger, 2004, for instance). This necessarily in-

[2]Another example for a model where decisions are shaped by distributional properties of the available allocations is the quasi-maximin model by Charness and Rabin (2002) which adds to material self-interest surplus maximization and the Rawlsian maximin motive as drivers for behaviour.

volves second-order beliefs entering the picture. Models incorporating second-order beliefs provide quite sophisticated theories of reciprocity. Unfortunately, they often yield multiple equilibria even in quite simple games and finding these is often not trivial. To avoid such problems, a second approach, the "revealed intentions" approach, has been proposed by Cox and Sadiraj (2007) and Cox et al. (2008a). Here, a second mover's benevolence in a two-player two-stage game is a function of the relative kindness or unkindness of the first mover as revealed by the objective characteristics of his (observed) choices. The first mover's kindness, in turn, is determined by the relative generosity of the opportunity set implied by his choice relative to alternative opportunity sets he could have chosen instead.

The present paper contributes to the revealed intentions approach of conditional other-regarding preferences by exposing subjects in the lab to a large number of two-player two-stage games and by studying how second movers react to the opportunities of gains and losses for each player generated by the choice of the first mover. Specifically, we expose subjects in the lab to graphical representations of two-player two-stage games in which (i) the first mover has to choose between two budget sets, one containing a single allocation, the other containing several possible payoff allocations; and (ii) the second mover has to choose one of the available payoff allocations in the non-trivial budget set – provided the first mover has chosen it. By systematically varying the two budget sets available to the first mover, we investigate how opportunities of gains and losses for each player influence the second mover's benevolence towards the first mover. We find that the possibility of gains for the second mover (generosity) and the risk of losses for the first mover (vulnerability) are important drivers for second mover behaviour. On the other hand, the

possibility to mutually gain and an aversion against violating trust seem to be far less important motivators. We also find that second movers compare the actual choice of the first mover and the alternative choices that would have been available to him to allocations that involve equal material payoffs.

Compared to the existing literature on conditional other-regarding preferences the present paper makes three critical contributions: The first contribution is the introduction and implementation of an experimental design in which subjects are exposed to geometric representations of choice sets; this allows for the collection of a large number of observations per subject which facilitates statistical analysis at the level of the individual decision-maker. Regarding this contribution, the paper closest to ours is Fisman et al. (2007). Those authors are interested in *unconditional* other-regarding concerns. As a consequence, in their experiments there is only one player role – that of a dictator – and each dictator is exposed to 50 different decision problems, each graphically represented as a linear budget set from which the subject can choose.[3] Since our main research focus is on *conditional* other-regarding preferences we extend this approach by having two player roles – the role of a first mover and the role of a second mover; the first mover chooses among graphical representations of opportunity sets while the second mover makes a dictator decision within a given opportunity set similar to the one subjects are asked to make in Fisman et al. (2007). By varying the set of budget

[3]This is the baseline experiment in Fisman et al. (2007). In addition to this, the authors also investigate two alternative treatments: one has linear budget sets as the baseline but differs from the latter in that each dictator decision has now consequences for two other persons (i.e. budget sets are three-dimensional in this treatment); the other has two-dimensional budget sets as the baseline but differs from the latter in having allocations in the choice set that differ only in the material payoff of the recipient, or only in the material payoff of the dictator (i.e. budgets are step-shaped in this treatment).

sets available to the first mover we are able to investigate how the second mover's choice varies with the budget set actually chosen by the first mover and with the counterfactual alternative opportunity set the first mover could have chosen instead.

Our second innovation is the experimental investigation of the relative importance of different motives for behaviour of players in extensive form games. In this respect the papers closest to ours are Cox (2004) and Cox et al. (2007, 2008a, 2014). While Cox (2004) employs a triadic experimental design to disentangle the relative importance of conditional and unconditional other-regarding preferences for behaviour of second movers in the investment game, the present paper's main aim is to disentangle the relative importance of different basic motives for the conditional part of players' other-regarding preferences. Similar to Cox et al. (2007, 2008a), we suppose that the second mover in a two-player two-stage game cares about how the opportunity set chosen by the first mover compares to alternative opportunity sets the first mover could have chosen instead. However, while these papers compare opportunity sets in terms of generosity by the first mover towards the second mover and focus on reciprocity as possible motivation for the second mover, we look not only at the possible gains for both players but also at possible losses and look at a broader array of possible motivations. In this latter respect our paper is similar to Cox et al. (2014). However, in contrast to that work, we look not only on trust game constellations and we collect many observations per individual.[4] The latter feature of our experimental design allows us to estimate utility functions at the individual level in a within-subjects design while Cox et al. (2014) derive their results from

[4]As will become clear later, the treatments in Cox et al. (2014) are all located in area 11 of Figure 5.2 while we expose subjects to decision situations in each of the cells in the figure.

comparisons of aggregate data across treatments in a between-subjects design.

Our third innovation is the introduction of a silent social norm – the equal-split norm – into the revealed intentions approach. In this respect, our paper is related to previous work on the importance of the equality norm for economic behaviour – see Fehr and Schmidt (1999), Bolton and Ockenfels (2000) and Andreoni and Bernheim (2009), for instance. While Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) stress the importance of the equal-split norm for unconditional other-regarding preferences, we show that this norm is also crucial for our understanding of conditional other-regarding preferences. Conditional other-regarding preferences might also be relevant for behaviour in the experiments reported by Andreoni and Bernheim (2009). However, while Andreoni and Bernheim are interested in the impact of "audience effects" on behaviour, we are interested in situations where audience effects are unlikely to play a role.

The remainder of the paper is organized as follows: Section 5.2 presents our experimental design. It is followed by our conceptual framework in Section 5.3, which consists of a classification of choice characteristics, our model of other-regarding preferences, and predictions derived from the model. In Section 5.4, we report our data and estimate the parameters of our model. Section 5.5 discusses our findings and concludes.

5.2 Experimental design

Our workhorse is a two-stage game with two players. In the first stage, the first mover (FM, he) makes a binary decision – he chooses between a fixed allocation (consisting of a payoff for himself and a payoff for the

second mover) and an opportunity set containing several possible payoff allocations. In the second stage, the second mover (SM, she) chooses a fixed allocation from the opportunity set whenever the FM has chosen this option – otherwise she has no move.

Our design can be seen as a (generalization of a) hybrid between an *investment game* (à la Berg et al., 1995) where both players have rich choice sets (provided the FM has made a "trusting choice") and a *mini trust game* (à la McCabe et al., 2003) where both players have only a binary choice to make (provided the FM has made the "trusting choice"): In our design, the FM has a binary choice to make (it can be interpreted as a choice between transferring a given amount s to the SM and not transferring anything) while the SM has a richer choice set (in our design a choice between seven allocations provided the FM has transferred s).[5]

We are interested in how the SM reacts to the opportunities of gains and losses for both players generated by the FM's choice. To investigate this question we expose subjects to a large number of graphical representations of choice situations. Across choice situations, we systematically vary the set of opportunity sets available to the FM in the first stage. By doing so, we can investigate how a wide range of "intentions" revealed by the FM's choice affect the SM's benevolence in the second stage.

The experiment was conducted by pencil and paper with students from a large Australian university. The subjects in the experiment were recruited via the ORSEE software by Greiner (2015). After subjects read the instructions (they are contained in Appendix B), they were read aloud by an experimenter. Subjects answered a couple of control questions to

[5]Some of the games investigated by Charness and Rabin (2002) constitute special cases of our design. They found in these cases that the SM often reciprocated to the kindness of the FM (as revealed by his choice). Our design systematically varies the set of choices offered to the FM to investigate other potential factors driving the behaviour of the SM.

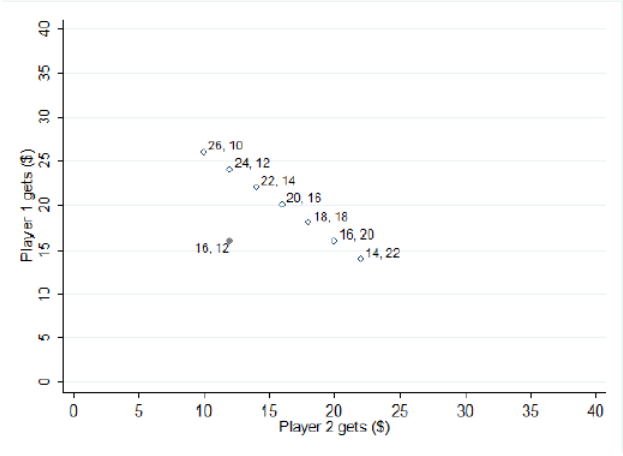
assure their understanding of the task and the payoff procedure. Then, each participant was randomly assigned a role, either the role of a FM or the role of a SM. The randomization was such that in each session we had the same number of FMs and SMs and the participants kept their roles during the entire session.

Subjects in both roles were faced with 60 graphical representations of sets of opportunity sets. Each set of opportunity sets consisted of two options, a fixed payoff allocation and the opportunity to let the SM make a decision among a set of seven possible payoff allocations threaded on a downward sloping straight line. In the following we call the former option *the point* and the latter *the line*. If the FM chooses *the point*, the SM has no further move while for *the line*, she has to decide among the allocations in the non-trivial opportunity set. To obtain all data from SMs we used the strategy method: Each SM was asked to make a decision as if the paired FM had assigned her the opportunity to make the choice between the seven payoff allocations on *the line*.[6]

Figure 5.1 shows three typical examples of a decision situation. The task of the FM (Player 1) is to check one of the boxes below the figure indicating whether he prefers option A, *the point*, or option B, *the line* of hollow dots. The task of the SM (Player 2) is to indicate her choice by circling her preferred allocation on *the line* of hollow dots. The 60

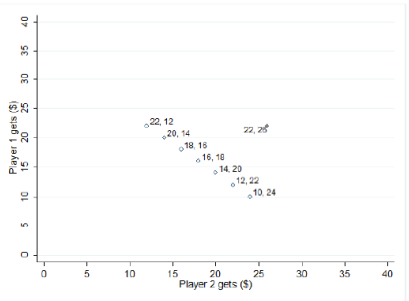
[6]The strategy method offers the benefit of making the responses in all decision nodes observable. While there are potential effects of using the strategy method instead of the direct-response method (such as a reduction in incentives or a "hot" vs. "cold" effect that might affect the participants' choices – see Zizzo, 2010, for a discussion), the experimental literature reports no case in which a treatment effect was observed with the strategy method and not with the direct-response method (see Brandts and Charness, 2011). We further argue that while it is likely that the strategy method has an impact on the level of the reaction strength of SMs, it is unlikely to have a systematic impact on their response to changes in the dimensions we are mainly interested in. Our analysis will focus on *changes* in the pro-sociality of responses rather than on the *level* of pro-social behaviour and thus the strategy method should most likely be innocuous (Charness and Levine, 2007).

decision tasks differed in the positions of the available opportunity sets and the positions were allocated randomly to pairs of subjects. The randomization ensured that *lines* stayed in the positive orthant and the location of *the point* was varied around the *lines* as depicted in Figure 5.2.[7]



Player 1: Please tick your preferred option
☐ Option A: The filled dot
☐ Option B: The line of hollow dots (Let Player 2 make a choice)

Player 2: Please choose one of the allocations on the line of hollow dots by circling it in the figure.



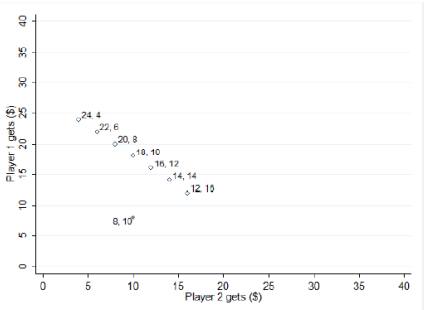


Figure 5.1: Typical decision tasks.

[7]The randomization also limits concerns for an indirect experimenter effect whereby participants observing systematic variations of a *point* location relative to the same *line* would infer that their behaviour is expected to change as a consequence of the relative position of *the point*.

In each session, one subject in the role of the FM and one subject in the role of the SM received exactly identical experimental questionnaires – that is, two experimental subjects in each session faced exactly the same 60 decision tasks. At the end of the experiment, we paired the subjects who received identical questionnaires. In each pair, we then picked randomly one of the 60 decision tasks, and paid the participants the payoffs corresponding to their joint choices in this situation. Overall, sessions lasted around one hour and participants earned $16.5, on average, plus a show up fee of $5. All monetary amounts are in Australian dollars.

5.3 Conceptual framework

5.3.1 Second mover's social preferences

In line with the revealed intentions approach, we suppose that the SM cares about how the opportunity set chosen by the FM compares to the alternative opportunity set the FM could have chosen instead. Similar to Cox et al. (2008a), we look at the possible gains for both players resulting from the FM's choice. Similar to the companion paper Cox et al. (2014), we extend this approach by also looking at the possible losses for both players. Compared to Cox et al. (2014), we study a richer array of possible motivations covering all constellations displayed in Figure 5.2.[8] We discuss the features of the areas in this figure in the

[8]The Cox et al. (2014) design comprises five treatments implemented between subjects. In all these treatments, the SM decides how to divide 60 experimental currency units between herself and the FM in case the FM sends her his endowment of 15. The treatments differ in what happens in case the FM decides not to send the endowment to the SM, and whether the FM can make such a decision at all. Thus, in the language of the current paper, the Cox et al. (2014) design keeps the location of the *line* constant and varies the location of the *point* and whether a *point* is available

next subsection.

To allow for errors in decision-making, we adopt a random utility approach. In our experiment, in each of the 60 decision tasks, the SM's opportunity set consists of seven discrete options. We therefore use a random utility discrete choice framework – see Train (2009) for details.

Experimental data from dictator games suggests that the egocentric altruism model by Cox et al. (2007) or a similar constant elasticity of substitution utility function represents revealed preferences quite well (see Andreoni and Miller, 2002, or Cox and Sadiraj, 2012, for instance). To incorporate reciprocal motivations, Cox et al. (2007) extend the egocentric altruism model by allowing an agent's willingness to pay for increases or decreases in the payoff of another person (hereafter "benevolence") to depend on this other person's prior actions (on whether the other person was kind or harmful to the agent). Specifically, Cox et al. (2007) propose a model where a subject's benevolence depends on his emotional state, which in turn depends on the other player's choice. For the two-player case the proposed utility function reads:

$$u(x_s, x_o) = \begin{cases} \left(x_s^\alpha + \theta x_o^\alpha\right) \alpha^{-1} & \alpha \in (-\infty, 0) \cup (0, 1], \\ \\ x_s x_o^\theta & \alpha = 0, \end{cases}$$

where x_s is the subject's own material payoff which contributes positively to his utility, x_o is the payoff of the other subject and α and θ are parameters, both supposed to be (weakly) smaller than one. The convexity is captured by α through the elasticity of substitution $\sigma = 1/(1-\alpha)$. The parameter θ is called the agent's "emotional state" and the effect

at all. In terms of Figure 5.2, the Cox et al. design only investigates constellations in area 11 while we expose subjects to decision situations in each of the cells in the figure.

of the other's payoff on utility depends on the sign of θ. A positive θ means that the individual under consideration cares positively for the other agent in the sense that he is willing to give up money to increase the other's payoff. The agent's willingness to pay – which is the amount of own income the agent is willing to give up in order to increase the other agent's income by one unit – is given by:[9]

$$WTP = \frac{\delta u/\delta x_o}{\delta u/\delta x_s} = \theta \left(\frac{x_s}{x_o}\right)^{1-\alpha}.$$

As is easily seen, the larger θ, the higher the WTP. Note further that α measures the importance of relative payoffs. For positive θ, $\alpha = 1$ yields linear preferences implying that the WTP is independent of relative payoffs, while $\alpha < 1$ yields convex preferences implying that the WTP and with it the agent's benevolence towards the other agent increases as the other's relative payoff decreases. And the more convex the preferences (the smaller α), the higher is the sensitivity of the WTP to changes in the relative payoff (x_o/x_s).

Here we adopt this functional form and – in line with Cox et al. (2007) – we capture intention-based benevolence from the SM by allowing her emotional state θ to depend on the FM's previous choice. Specifically, we allow a SM's θ to depend on the observable characteristics as defined in the next subsection:

$$\theta = \theta(\text{observable characteristics of FM' actual choice}).$$

[9]For interpretation purposes, it is easier to talk about the willingness to pay (WTP) than the more more familiar marginal rate of substitution (MRS), which is given by: $MRS = \frac{\delta u/\delta x_s}{\delta u/\delta x_o} = \theta^{-1}\left(\frac{x_o}{x_s}\right)^{1-\alpha} = 1/WTP$.

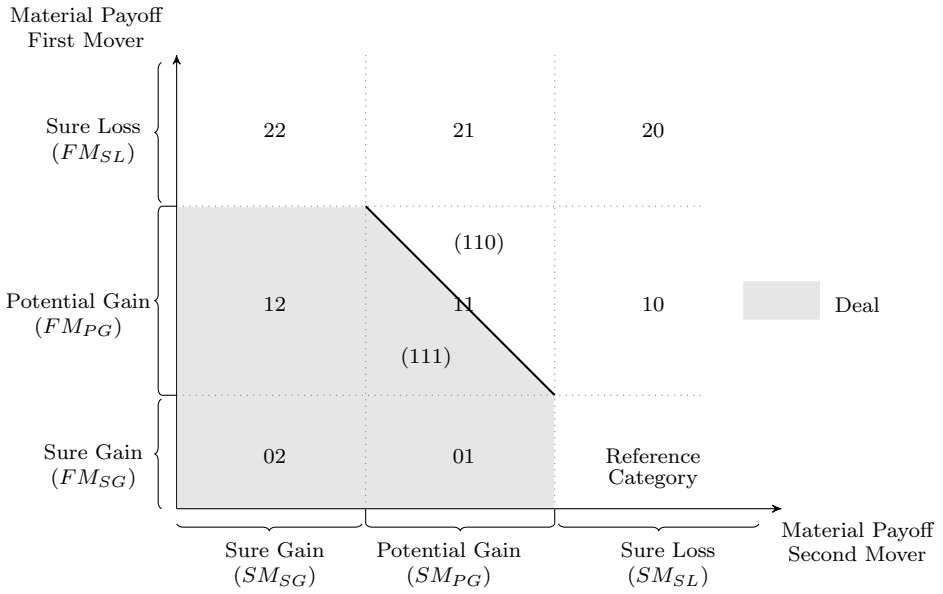


Figure 5.2: Observable characteristics of the FM's choice when choosing *the point* for different positions of *the point* relative to *the line*.

5.3.2 Classification of first mover's choices, attributed intentions and their impact on second mover's behaviour

The classification in Figure 5.2 is based on the gain/loss principle applied to both players, i.e. whether the opportunity set that was chosen by the FM (*the line*) comes with an actual or potential increase or decrease of each player's payoff compared to the not chosen opportunity set (*the point*). Our first hypothesis is motivated by the experimental evidence indicating that reciprocity is an important driver for behaviour in games. Reciprocation entails responding to positive perceived kindness with positive kindness, and to negative perceived kindness with negative kindness (Rabin, 1993; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004). In a material context, kindness is usually equated with generosity. To formulate a hypothesis regarding the impact of positive reciprocity on

the behaviour of the SM, we therefore characterize the choice of the FM in terms of the implied generosity towards the SM. Here, we distinguish between three levels of *generosity* when the FM chooses *the line* over *the point*:

Definition 5.1 *Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses the line.*

a) If the chosen opportunity set (the line) only includes allocations which decrease the SM's payoff compared to the not chosen opportunity set (the point), the FM's choice (of the line) is said to imply a ***sure loss for the SM****.*

b) If the line includes allocations for which the SM's payoff is (weakly) higher, and allocations for which the SM's payoff is (weakly) lower compared to her payoff in the point, the FM's choice of the line is said to imply a ***potential gain for the SM****.*

c) If the line only includes allocations which increase the SM's payoff compared to the point, the FM's choice of the line is said to imply a ***sure gain for the SM****.*

Using this classification of FM behaviour, it seems plausible that choices of the FM that imply a sure gain for the SM are interpreted by the SM as more generous than choices that imply a potential gain for the SM, and that choices that imply a potential gain for the SM are interpreted as more generous than choices that imply a sure loss for the SM. This consideration yields our first prediction:

Hypothesis 5.1 ***(Impact of Generosity)***
The SM's benevolence increases with the level of generosity implied by the choice of the FM. That is, the SM becomes progressively more benevolent when we move from situations where the FM's choice implies a sure loss for the SM, to situations where the FM's choice implies a potential gain for the SM, to situations where the FM's choice implies a sure gain for the SM.

Our second hypothesis is based on experimental evidence indicating that the vulnerability of the FM is an important driver for the behaviour of the SM in the investment game (see Cox et al., 2014, for an investigation of the role of vulnerability in the investment game). Vulnerability in our context means that the FM, by choosing *the line*, accepts the risk of losing money depending on the SM's choice. To formulate a hypothesis regarding the impact of vulnerability on the behaviour of the SM we therefore characterize the choice of the FM in terms of the implied risk for the FM. Here, we distinguish between three levels of *vulnerability* of the FM when he chooses *the line* over *the point*:

Definition 5.2 *Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses the line.*

a) If the chosen opportunity set (the line) assures the FM a payoff increase compared to the not chosen opportunity set (the point), the FM's choice is said to imply a ***sure gain for the FM****.*

b) If the line includes allocations for which the FM's payoff is (weakly) higher, and allocations for which the FM's payoff is (weakly) lower compared to the payoff in the point, the FM's choice of the line is

*said to imply a **potential gain for the FM**. In that case, we also say that the FM's choice of the line makes him **vulnerable**.*

c) *If the line only includes allocations which decrease the FM's payoff compared to the point, the FM's choice of the line is said to imply a **sure loss for the FM**. In this case, we also say that the FM's choice of the line corresponds to a **sacrifice**.*

Using this classification of FM behaviour, we now posit two hypotheses. Hypothesis 5.2a predicts that choices by the FM that make him vulnerable lead to benevolent behaviour by the SM:

Hypothesis 5.2a ***(Impact of Vulnerability)***
The SM's benevolence increases if the FM's choice of the line makes him vulnerable. Specifically, the SM becomes more benevolent when we move from situations where the FM's choice of the line implies a sure gain for the FM, to situations where the FM's choice of the line implies a potential gain for the FM.

We also suspect that FM choices that correspond to a sacrifice influence the behaviour of the SM. This is the content of Hypothesis 5.2b. Note that Hypothesis 5.2b does not make any prediction on how the effect of sacrifice compares to the effect of vulnerability.

Hypothesis 5.2b ***(Impact of Sacrifice)***
The SM's benevolence increases if the FM's choice of the line implies a sacrifice for him. Specifically, the SM becomes more benevolent when we move from situations where the FM's choice of the line implies a sure gain for the FM, to situations where the FM's choice of the line implies a sure loss for the FM.

Our next hypothesis is based on the idea that SMs may reward FM choices that have the potential to increase the payoffs of both parties. This conjecture is motivated by the experimental evidence indicating that efficiency concerns are important for behaviour in the lab and in the field (see Engelmann and Strobel, 2004; Fehr et al., 2006, among others). To formulate a hypothesis regarding the impact of concerns related to Pareto-efficiency on SM behaviour, we characterize FM choices according to the payoff consequences for both players as follows:

Definition 5.3 *Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses the line. If the line includes allocations which represent a Pareto improvement relative to the point, the FM's choice of the line is said to allow for a* ***deal****.*

We then state:

Hypothesis 5.3 ***(Impact of Deal)***
The SM's benevolence increases if the FM's choice of the line allows for a deal. That is, the SM becomes more benevolent when we move from situations where the choice of the FM does not allow for a Pareto improvement to situations that allow for a mutual improvement.

Our next (and last) hypothesis is motivated by the large experimental literature on trust and trustworthiness. In experimental economics, the most frequently used instrument to study the importance of those concepts for behaviour is the investment game (Berg et al., 1995) and its close relative, the binary trust game (studied by McCabe *et al.*, 2003, for instance). There is by now an impressing amount of evidence indicating that SM behaviour in those games is neither consistent with own-money

maximization nor in line with purely distributional concerns (see Cox, 2004; Ashraf et al., 2006; Chaudhuri and Gangadharan, 2007; Cox et al., 2008b, 2014, among others). Less clear is the answer to the question what is really driving SM behaviour in this class of games. Here, we address this question indirectly by investigating whether FM behaviour characterized by the combination of characteristics defining a trusting move in the investment game induces more benevolence in the SM than behaviour characterized by other combinations. To formulate a hypothesis regarding the impact of trusting acts by the first mover on the behaviour of the second mover we define:

Definition 5.4 *Let the FM choose the actual opportunity set for the SM from a collection consisting of a point and a line. Suppose the FM chooses the line. If the choice of the line makes the FM vulnerable and if in addition it allows for a deal, then the FM's choice is said to reveal* ***trust****.*

We then hypothesize that choices revealing trust have the power to trigger benevolence in the SM:

Hypothesis 5.4 ***(Impact of Trust)***
The SM's benevolence increases if the FM's choice of the line reveals trust. That is, the SM becomes more benevolent when we move from situations where the choice of the FM does not reveal trust to situations where the choice of the FM reveals trust.

5.4 Data and results

We first provide an overview of the data collected in our experiment and a descriptive analysis. We then proceed with the parameter estimation

of our model.

5.4.1 Data

We carried out 14 experimental sessions involving 190 subjects in total. Since our research focus lies on the conditional part of an individual's social preferences, we are only interested in the data collected from experimental SMs. Since we collected the data via the strategy method, our data set consists of 60 decisions for each of the 95 SMs.

Looking at the individual data, we find that 37 subjects (that is 38.9 percent of our SM population) behaved in a perfectly selfish way by choosing the lowermost point on *the line* in each of the 60 decision situations. Hence, $\theta = 0$ and $u(x_s, x_o) = x_s$ for almost 40 percent of our SM sample. This compares to previous studies (Andreoni and Miller, 2002; Fehr and Gächter, 2000), where typically completely selfish behaviour was reported for between 20 and 50 percent of individuals.

Nevertheless, the majority of our subjects (61.1 percent) behaved in a way that is inconsistent with the only self-interested rationalist assumption by not choosing the selfish allocation in 57.6 percent. We therefore observe that many participants exhibit other-regarding preferences of some kind.

For our further analyses, we exclude the purely selfishly acting SMs from our data sample and focus on the 58 participants that reveal some form of other-regarding behaviour.[10] The overall distribution of the choices of those SMs is presented in Figure 5.3 and Table 5.1.[11] In Table 5.1, we see that the uppermost four points on *the line* (points 4-7) are

[10]Since $\theta = 0$ for purely selfishly acting individuals, the behaviour of subjects in this subsample is not informative about how intentions influence social preferences.

[11]The experiment was conducted by pen and paper and a small number of answers (N=17) were missing in the questionnaires. This leaves a dataset of 3463 observations.

chosen in only 27.6 percent of decision tasks. This is not really surprising as point 7 is the most benevolent decision a SM can make, and point 1 is the least benevolent one. Thus, the subjects in the subsample under consideration – although not purely selfish – have a tendency to care more for their own than for the other's payoff.

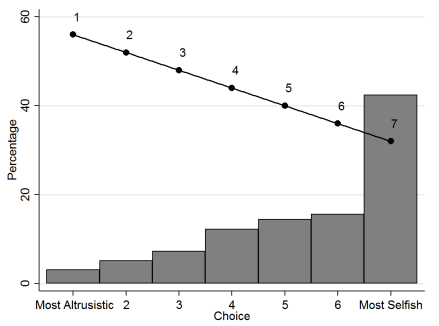


Figure 5.3: SMs' choice distribution.

Choice	Freq.	%	Cum. %
7 (Most Altruistic)	106	3.1	3.1
6	177	5.1	8.2
5	252	7.3	15.4
4	422	12.2	27.6
3	498	14.4	42.0
2	540	15.6	57.6
1 (Least Altruistic)	1468	42.4	100.0
Total	3463	100.0	

Table 5.1: Summary of participants' choices.

5.4.2 Descriptive analysis

In a first step, we analyse whether the characteristics defined in Section 5.3 influence SM behaviour. For this purpose, we define a set of binary variables reflecting Definitions 5.1 and 5.2 introduced in Subsection 5.3.2: "sure gain for the FM" (FM_{SG}), "potential gain for the FM" (FM_{PG}) and "sure loss for the FM" (FM_{SL}), as well as "sure gain for the SM" (SM_{SG}), "potential gain for the SM" (SM_{PG}) and "sure loss for the SM" (SM_{SL}). In addition, we analyse the effect of the dummy "*Deal*", which is one if the choice of *the line* allows for a deal according to Definition 5.3 and zero otherwise; we also analyse the effect of the dummy "$Trust$", which is one if the choice of *the line* reveals trust according to Definition 5.4 and zero otherwise. Note that the shaded areas in Figure 5.2 cover

situations where the choice of the line allows for a deal, while area 111 contains all situations where the choice of the line reveals trust.

Our first observation supports our main hypothesis that the choice of the SM on *the line* depends significantly on the nature of the counterfactual choice the FM could have made: Figure 5.4 displays the mean SM choice as a function of the characteristics of the FM's choice. The significance of the difference in means is indicated using t-tests (from regressions on dummies using cluster-robust variance to control for the non-independence of observed choices within participants). The data displayed in Figure 5.4 suggests that the choices of SMs become more benevolent if the level of generosity increases from SM_{SL} to SM_{PG} ($p = 0.072$) and from SM_{PG} to SM_{SG} ($p = 0.010$). SMs seem also to become more benevolent if the FM's choice implies vulnerability – moving from FM_{SG} to FM_{PG} ($p < 0.001$) – or sacrifice – moving from FM_{SG} to FM_{SL} ($p = 0.012$). Interestingly, the mean choice of SMs is not significantly different between situations characterized by FM_{PG} and situations characterized by FM_{SL} ($p = 0.437$). Turning to *Deal* and $Trust$, we find that SMs are relatively more benevolent when the choice of the FM allows for a *Deal* ($p = 0.036$) or reveals $Trust$ ($p = 0.027$). It should be noted however that these latter observations do not imply that SMs react to *Deal* and $Trust$ per se; they might rather react to the FM's generosity and vulnerability which are both present in situations of *Deal* and $Trust$.

The effect of the counterfactual choice the FM could have made on SM behaviour can also be seen in Figure 5.5. In this figure the cumulative distribution functions (CDFs) of SM choices on *the line* are represented depending on the level of generosity, the level of vulnerability, and on whether the choice of the FM allows for a *Deal* or reveals $Trust$. A first-order stochastically dominating CDF reflects more benev-

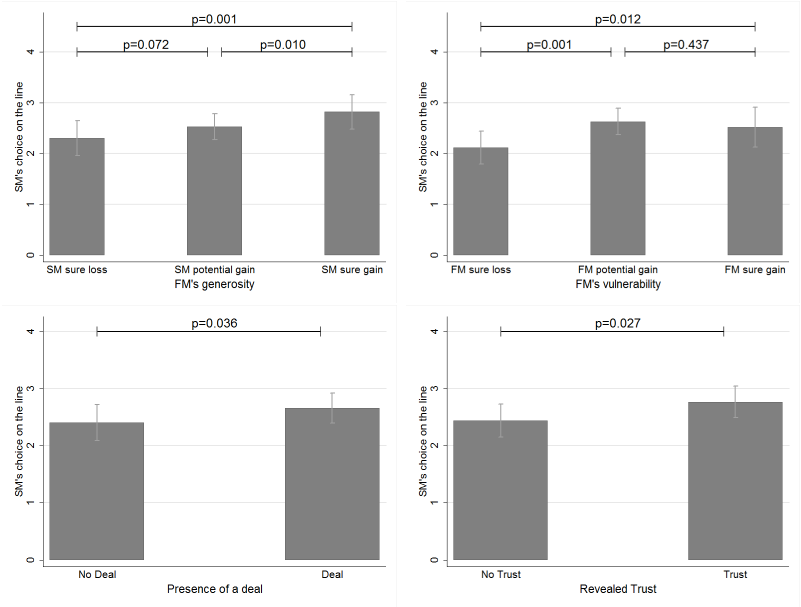


Figure 5.4: SM's benevolence as a function of the characteristics of the FM's choice. The figure shows the average choice of the SMs on the line. Higher values indicate more benevolence.

olence. It can be seen that the CDF for choices exhibiting SM_{SG} first-order stochastically dominates the CDF for choices featuring SM_{PG} (KS test: $p = 0.025$), which in turn first-order stochastically dominates the CDF for FM choices featuring SM_{SL} (KS test: $p = 0.005$). This finding strengthens the previous result that a more generous choice by the FM triggers a more benevolent response by the SM, and therewith provides further support for Hypothesis 5.1. We also find support for Hypothesis 5.2. The CDF of choices featuring FM_{PG} first-order stochastically dominates the CDF of choices with FM_{SG} (KS test: $p = 0.003$), which is clearly in line with Hypothesis 5.2a. It is also the case that the CDF of choices featuring FM_{SL} first-order stochastically dominates the CDF of choices with FM_{SG} (KS test: $p = 0.034$), which is in line with Hypothesis 5.2b. Comparing the distribution of choices featuring FM_{PG} to the distribution of choices featuring FM_{SL}, we see that they differ (KS test:

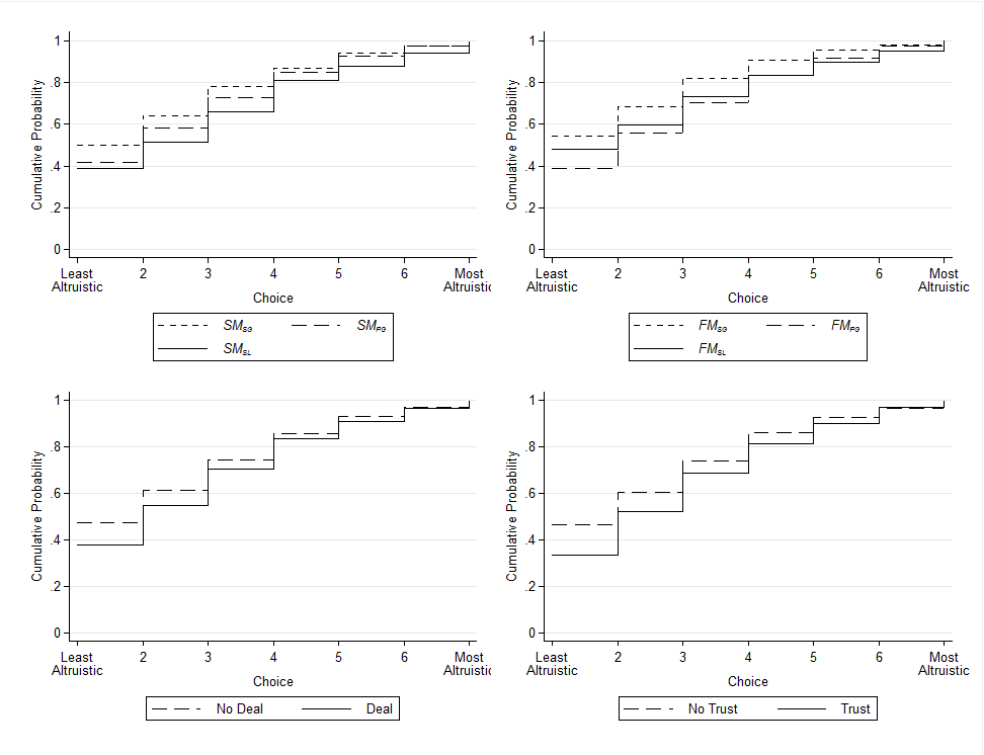


Figure 5.5: Cumulative distributions of the SM's choice by characteristics of the FM's choice.

$p = 0.001$) although the mean choice is statistically indistinguishable between the two situations. Specifically, the distribution of SMs' responses to FM_{SL} features both more most altruistic choices and more least altruistic choices. In the second row of Figure 5.5, we see that the CDF of choices that allow for a *Deal* first-order stochastically dominates the CDF of choices without a *Deal* available (KS test: $p < 0.001$). However, as previously stated this finding might be confounded by the fact that if the FM's choice allows for a *Deal*, it necessarily also entails either SM_{SG} or SM_{PG} which might be responsible for the effect on SM's benevolence. Finally, we also find some support for Hypothesis 5.4: The CDF of SM choices featuring $Trust$ almost first-order stochastically dominates the CDF of SM choices not revealing $Trust$ (KS test: $p < 0.001$). Again,

this finding might be confounded by the fact that the SM may simply react to the generosity and vulnerability which are present in the trust situation.

5.4.3 Disentangling revealed intentions

The structural model described in Subsection 5.3.1 makes it possible to disentangle the effects of different characteristics of the FM's choice on the SM's behaviour. Following the random utility approach (Train, 2009), we assume that the utility of SM i for payoff pair $x = (x_s, x_o)$ in a choice situation featuring the characteristic combination j includes a stochastic term which represents the unobserved part of utility (including quixotic variations in utility due to cognitive limitations when assessing the options):

$$v_j^i(x) = \left(x_{SM}^{\alpha} + \theta_j^i x_{FM}^{\alpha}\right) \alpha^{-1} + \varepsilon, \tag{5.1}$$

with

$$\begin{aligned} \theta_j^i \;=\; & \theta_{00} + \theta^i + \\ & + \beta_{FM_{PG}} \mathbb{1}_{j\in\{10,110,111,12\}} + \beta_{FM_{SL}} \mathbb{1}_{j\in\{20,21,22\}} + \\ & + \beta_{SM_{PG}} \mathbb{1}_{j\in\{01,110,111,21\}} + \beta_{SM_{SG}} \mathbb{1}_{j\in\{02,12,22\}} + \\ & + \beta_D \mathbb{1}_{j\in\{01,02,111,12\}} + \beta_T \mathbb{1}_{j=111}. \end{aligned} \tag{5.2}$$

Here, θ_j^i is the emotional state of SM i when she observes that the FM has chosen *the line* in a choice situation where the alternative choice he could have made (that is, *the point*) is located in area $j \in \{01, 02, 10, 110, 111, 12, 20, 21, 22\}$ as defined in Figure 5.2. This formulation assumes that "motives are additive" in the sense that adding a given motive has the same effect independently of whether other motives are present or absent.

We will relax this assumption later on. Note also that this formulation allows for individual heterogeneity in social preferences with the inclusion of an individual specific term θ^i. We follow a standard approach in discrete choice modeling (Train, 2009) in assuming $\varepsilon \rightsquigarrow Gumbel(\lambda)$. This implies that the choice model is a non-linear multinomial logit model with the probability that a given allocation x' is chosen among a set X of possible allocations given by:

$$\mathbb{P}(x') = \frac{\exp(\lambda v(x'))}{\sum_{x \in X} \exp(\lambda v(x))},$$

where λ is the subjects' precision parameter.[12] We estimate the parameters α and θ^i_j by maximum-likelihood. Each participant provided 60 data points, we therefore cluster the standard error by participant.

Table 5.2 reports the estimates of our basic model. As expected $\alpha < 1$ which indicates convex preferences. In the sequel, we focus our discussion on the parameter θ_j since this is the parameter related to our research question. The impact of the characteristics of the FM's choice on this parameter is measured in comparison to the reference categories FM_{SG} and SM_{SL}. These reference categories are arguably associated with the lowest level of benevolence by the SM.

The parameter estimates in Table 5.2 suggest that – starting from the reference categories – an increase in the level of generosity from the FM towards the SM, as well as an increase in the FM's vulnerability indeed have a significant positive impact on the SM's altruism coefficient θ and thus on her benevolence. Regarding generosity, we find that a sure gain for the SM (SM_{SG}) has a significant effect on the SM's benevolence while a sheer potential gain (SM_{PG}) does not have a significant effect. This

[12] This is called the "Luce model" (see Wilcox, 2008).

Model (N=3463):		$v_j^i(x) = \left(x_{SM}^\alpha + \theta_j^i x_{FM}^\alpha\right)\alpha^{-1} + \varepsilon$	
Parameter		Estimate	Robust SE
α		0.282	0.170
θ	**FM payoffs**		
	FM_{SL}	0.188*	0.080
	FM_{PG}	0.190**	0.071
	FM_{SG}	(ref)	
	SM payoffs		
	SM_{SG}	0.206*	0.098
	SM_{PG}	0.042	0.038
	SM_{SL}	(ref)	
	$Deal$	0.001	0.066
	$Trust$	0.027	0.075
λ		4.078**	1.391

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.2: Estimation of α and θ_j^i by maximum-likelihood taking FM_{SG} and SM_{SL} as reference categories.

result provides partial support for Hypothesis 5.1:

Result 5.1 ***(Impact of Generosity)***

The SM's altruism coefficient θ and therewith her benevolence increases with the level of generosity implied by the choice of the FM. However, the effect is significant only for situations where the FM's choice implies a sure gain for the SM.

Turning to the effect of vulnerability, we see that FM_{PG} raises θ significantly. This result confirms the finding of the descriptive analysis and is in line with Hypothesis 5.2a. In line with Hypothesis 5.2b, we also find a positive effect of FM_{SL} on the SM's benevolence. Comparing the two, we see that the estimated coefficients of FM_{SL} and FM_{PG} are roughly equal. Thus, acts that make the FM vulnerable and acts that imply a sure loss for the FM seem to have a similar impact on the intention-perception of the SM as revealed by her behaviour.

Result 5.2 ***(Impact of Vulnerability and Sacrifice)***

The SM's altruism coefficient θ and therewith her benevolence increases if the choice of the FM entails vulnerability (potential loss for the FM) or sacrifice (sure loss for the FM). Comparing the two effects, we see that they are similar in size.

Investigating the question whether the behaviour of SMs becomes more benevolent when the choice by the FM allows for a Pareto improvement, we observe that *Deal* availability has no significant effect on the benevolence of the SM. Hypothesis 5.3 is therefore not supported by the data. The previously observed shift in the CDF of SM's choices (Figure 5.5) seems indeed to be driven by generosity and/or vulnerability.

Result 5.3 ***(Impact of Deal)***

The availability of a deal by itself has no effect on the second mover's altruism coefficient θ and therewith on her benevolence.

Similarly, we do not observe any effect of trust in itself when the potential gains and losses of the two players are controlled for. Hypothesis 5.4 is therefore not supported by the data either. Here again, the shift in the CDF of the SM's choices between situations where the FM's choice reveals trust and situations where it does not (Figure 5.5) seems to be driven by the effects of generosity and vulnerability without an additional impact of trust in itself.

Result 5.4 ***(Impact of Trust)***

The expression of trust has no effect in itself on the SM's altruism coefficient θ and therewith on her benevolence.

As previously mentioned our estimation of Model (5.1) assumes that motives are additive in Equation (5.2). We now relax the additivity

assumption and allow for possible interactions between the FM's vulnerability and his generosity towards the SM. Specifically, we define a dummy for each area displayed in Figure 5.2 and estimate the model:

$$\theta_j^i = \theta_{00} + \theta^i + \sum_k \beta_k \mathbb{1}_{j=k},$$

with $j, k \in \{01, 02, 10, 110, 111, 12, 20, 21, 22\}$.

Our chosen reference category is $FM_{SG} \times SM_{SL}$ (area 00 in Figure 5.2). The estimation results of this model are presented in Figure 5.6. By and large the results confirm our earlier findings. We observe that the SM's benevolence is high in situations where the FM's choice makes him vulnerable as long as vulnerability comes together with either a potential or a sure gain for the SM (SM_{SG}, SM_{PG}). The SM's benevolence is also always significantly positive for situations where the FM's choice implies a sacrifice, and, as long as the choice implies either a potential or a sure gain for the SM, there is no significant difference between the reaction of the SM to vulnerability and her reaction to sacrifice (no significant differences between β_{12} and β_{22} and between β_{110}, β_{111} and β_{21}). When the FM's choice implies a sure loss for the SM (SM_{SL}), the SM's benevolence increases with the opportunities of losses for the FM: FM_{PG} has a positive but insignificant effect and FM_{SL} has a positive and significant effect. This latter effect seems rather strange at first sight and it is investigated further in the next subsection.

While SM_{PG} and FM_{PG} in isolation are not enough to influence the benevolence of the SM (β_{01} and β_{10} are not significantly different from zero), it is noteworthy that their joint presence (in β_{110} and in β_{111}) does. It therefore looks like there is an interaction between the effect of generosity and vulnerability. Increasing the level of generosity to SM_{SG}

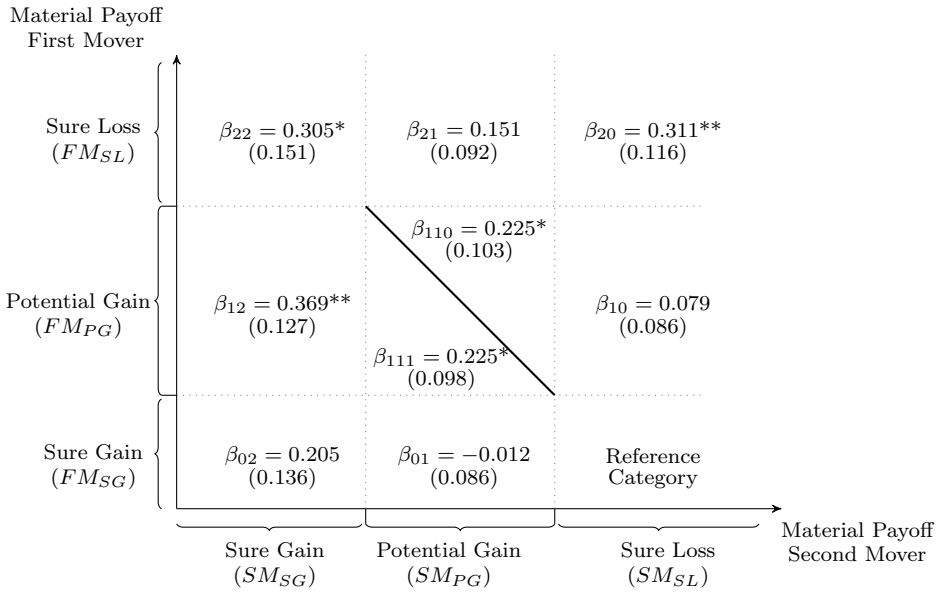


Figure 5.6: Maximum-likelihood estimation results of θ^i_j. The chosen reference category is $FM_{SG} \times SM_{SL}$. Robust standard errors are displayed in brackets. Significance: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

enhances the SM's benevolence even further (β_{12} is significantly larger than β_{111} and β_{110}), assuring the highest level of benevolence by the SM observed in our experiment.

Turning to Hypotheses 5.3 and 5.4 about the impact of *Deal* and *Trust* on the behaviour of the SM, we see that the coefficients β_{110} and β_{111} do not significantly differ from each other. Area 111 corresponds to FM choices revealing *Trust* and it differs from area 110 by the presence of a *Deal*. Thus, the relatively high level of benevolence from the SM observed in the area 111 seems solely be driven by the presence of FM_{PG} and SM_{PG}. This finding supports and strengthens our previously stated Results 5.3 and 5.4.

Overall, we conclude that relaxing the assumption that motives are additive does not change our previous results qualitatively: Positive reciprocity – whereby a generous choice by the FM triggers a benevolent re-

sponse by the SM – and vulnerability-responsiveness – whereby a choice by the FM that exposes him to the risk of losing money triggers a benevolent response – seem to be important drivers for SM behaviour, while deal-responsiveness – where the SM reacts positively to choices that create the possibility of mutual improvements – or trust-responsiveness – where the SM rewards acts that reveal trust – seem behaviourally less relevant.

5.4.4 Interpreting intentions from observed actions and salient social norms

In the precedent analyses, we have investigated whether a SM's benevolence is affected by the objective characteristics of the FM's choice – specifically by how his actual choice compares to the counterfactual alternative choice he could have made instead. By doing so, we have extended the revealed intention approach and looked at the possible gains and losses created by the FM's decision. Here, we argue that this approach can be extended further by incorporating the possible role of pre-existing *social norms* in the analysis. Social norms are by definition shared and common knowledge (Krupka and Weber, 2013). In games where allocations of resources are made between players, prevailing social norms may point to a "fair" allocation, that is one which would be considered as such by the different players. In an experiment where subjects enter the laboratory as equals, where they are allocated randomly to their roles, and where the money to be divided is a windfall provided by the experimenter, it seems plausible that fairness norms point to an equal split. Even though equal sharing might not be the only norm prevalent in the population of experimental subjects (e.g. asymmetry of roles may

be considered as giving different entitlements to different players), it is likely to be the most prevalent norm among all possible splits.

A look at the choices of experimental SMs suggests that the equality norm has indeed an impact. Figure 5.7 shows by how much the SM's choice differs from the least unequal allocation (the feasible allocation on *the line* that is closest to the 45 degree line, henceforth LUA).[13] Positive (negative) entries in Figure 5.7 correspond to choices on*the line* where the SM earns more (less) in material terms than the associated FM. As can be seen from the figure, there is a large concentration of SM choices at the LUA (more than a third of all choices by experimental SMs are at the LUA) and there is a pronounced discontinuity in the distribution of choices immediately to the left of the LUA, arguably because there is no social norm that dictates to give more than the "fair share" (implied by the LUA) to the other player. It therefore seems that the 50-50 split indeed plays a role for SM behaviour.

We next ask whether the interpretation of the FM's intentions by the SM is influenced by this norm. To address this question, we extend the revealed intentions approach by investigating whether choices are affected by the fairness of the counterfactual choice, *the point*, taking equality as the yardstick. Specifically, we estimate θ_j^i, using Equation (5.1), separately for situations where *the point* is above the 45 degree line and situations where it is below that line. Table 5.3 displays the associated parameters. It shows that in both subsamples coefficients have the same sign as reported for the aggregate data, but the parameters are

[13] If *the line* crosses the 45 degree line and if the crossing point is one of the seven feasible allocations on*the line*, then this allocation is the LUA; if *the line* crosses the 45 degree line but the crossing point is not a feasible allocation then the feasible allocation on *the line* that is closest to the 45 degree line is the LUA; and if*the line* does not cross the 45 degree line then the feasible allocation on*the line* that is closest to the 45 degree line is the LUA.

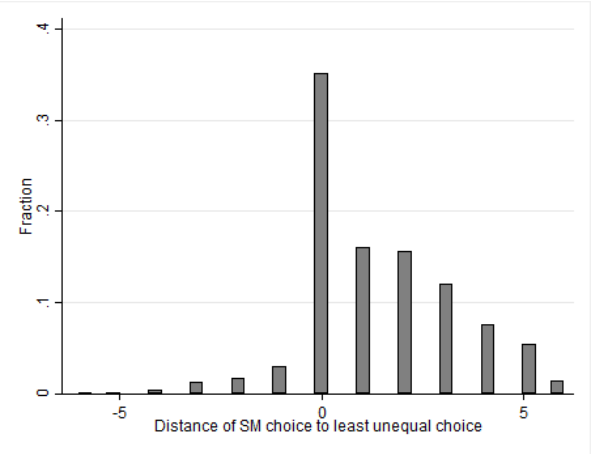


Figure 5.7: Distribution of the distance between the actual choice of the SM and the least unequal allocation on the line. Positive numbers represent unequal choices in favour of the SM, negative numbers represent unequal choices in favour of the FM.

smaller and not significant when *the point* is an allocation that favours the FM, while the coefficients of FM_{SL} and SM_{SL} are relatively large and significant when *the point* is to the advantage of the SM. Overall, this result suggests that intentions are read in relation to the 50-50 social norm. The SM reacts more positively to the generosity of the FM and to his vulnerability when the FM chooses *the line* in a situation where *the point* is an allocation characterized by inequality in favour of the SM. In such situations, by being generous, the FM is offering potential gains to the SM even though the SM was already advantaged by the initial allocation. By making himself vulnerable, the FM gives the possibility to the SM to make the FM worse off, even though the FM was already disadvantaged by the initial allocation. Therefore, one interpretation is that in such situations the generosity from the FM is perceived as particularly kind and the choice to make himself vulnerable particularly noticeable.

Turning to the result that the SM is relatively benevolent in the

Model: $v_j^i(x) = \left(x_{SM}^{\alpha} + \theta_j^i x_{FM}^{\alpha}\right)\alpha^{-1} + \varepsilon$

		Start favours FM		Start favours SM	
Parameter		Estimate	Robust SE	Estimate	Robust SE
α		0.681*	0.293	-0.135	0.199
θ	**FM payoffs**				
	FM_{SL}	0.032	0.089	0.181*	0.092
	FM_{PG}	0.070	0.110	0.134	0.082
	FM_{SG}	(ref)			
	SM payoffs				
	SM_{SG}	0.074	0.100	0.275*	0.118
	SM_{PG}	0.021	0.045	0.065	.049
	SM_{SL}	(ref)			
	$Deal$	-0.004	0.038	-0.054	0.092
	$Trust$	0.034	0.052	-0.058	0.086
λ		3.350**	0.893	12.245*	5.934
N		1832		1631	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.3: Estimation of α and θ_j^i by maximum-likelihood for situations where *the point* is above the 45 degree line and situations where it is below that line, taking FM_{SG} and SM_{SL} as reference categories.

$FM_{SL} \times SM_{SL}$ situation, we observe that the choice of *the line* by the FM in this constellation can potentially be interpreted as an attempt to avoid a split that is unfavourable to him, even if this leads to a loss in the payoffs of both players. To test whether this interpretation is consistent with the data, we re-estimate the Model (5.1) allowing for different values of the parameter θ in $FM_{SL} \times SM_{SL}$ situations above and below the diagonal. Table 5.4 displays the results. Column (1) allows θ to depend on a dummy $Point_{FM}$ taking a value 1 if *the point* favours the FM and zero if it favours the SM (we do not consider situations of equality). We find that the benevolence is overall larger for situations where the FM abandoned a relatively advantageous *point* when choosing *the line* ($p < 0.001$). In column (2), we interact this dummy with a dummy for the $FM_{SL} \times SM_{SL}$ situation. We find that SMs are significantly more benevolent when the $FM_{SL} \times SM_{SL}$ situation appears for *points* below the diagonal. This effect vanishes when the fixed allocation is above the diagonal.

These results are important for the revealed intentions approach. They show that the reaction of the SM to the FM's choice is not only shaped by differences between the opportunities generated by the choice set selected by the FM and the opportunities which could have been generated by a counterfactual choice. The SM's reaction seems also to depend on how a prevailing social norm of fairness labels each of these opportunities as fair or not. In the case of our experiment, the puzzling behaviour of the SM in $FM_{SL} \times SM_{SL}$ situations makes sense if the SM interprets the FM's choice as an attempt to avoid a split that is unfavourable to him. This as a consequence may induce the SM to make a more benevolent choice than in $FM_{SG} \times SM_{SL}$ situations. By contrast, benevolence by the SM is not observed when the (not chosen) point was favourable to

the FM.

Model (N=3203): $v_j^i(x) = \left(x_{SM}^{\alpha} + \theta_j^i x_{FM}^{\alpha}\right)\alpha^{-1} + \varepsilon$

Parameter		(1) Estimate	(1) Robust SE	(2) Estimate	(2) Robust SE
α		0.288	0.178	0.308	0.179
θ					
	$Point_{FM}$	0.170***	0.043	0.177***	0.044
	$\mathbb{1}_{FM_{SL} \times SM_{SL}}$			0.254***	0.089
	$Point_{FM} \times \mathbb{1}_{FM_{SL} \times SM_{SL}}$			-0.322*	0.153
λ		4.377**	1.497	4.241**	1.435

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5.4: Benevolence as a function of the position of the (not chosen) *point* relative to the equal-material payoff line.

5.5 Discussion

The empirical study of conditional other-regarding preferences based on higher-order beliefs is difficult because such beliefs are not observable and because eliciting them is a tricky task. An elegant alternative to belief-based conditional other-regarding preferences is the revealed intentions approach where a player cares about the generosity of the opportunity set chosen by another player compared to other opportunity sets that could have been chosen. The present paper has extended the revealed intentions approach by allowing agents to care not only about the possibility of gains generated by other agents' actions but also about the possibility of losses. In a two-player two-stage game, we have investigated how the second mover's other-regarding preferences are affected by different characteristics of the opportunity set chosen by the first mover compared to a counterfactual opportunity set the first mover could have chosen. By sys-

tematically varying the set of opportunity sets the first mover can choose from and investigating the response of the second mover to the actual choice of the first mover and the alternative choice he could have made, we were able to elicit how the second mover reacts to a wide variety of intentions as revealed by the first mover's choice.

We found that second movers do react to the possibilities of gains and losses generated for them and for the associated first mover. Second movers are typically more benevolent when the choice of the first mover creates an opportunity of gain for the second mover. This can be interpreted as a manifestation of positive reciprocity from the second mover. We have also seen that second movers react to the payoff consequences for the first mover implied by his choice with second movers becoming more benevolent when the first mover chooses an opportunity set that implies either a potential or a sure loss for him. These results suggest that future research should investigate further how self-imposed vulnerability and losses of a player affect the intention-based preferences of another player.

Our approach makes it possible to study complex "revealed preferences" as specific combinations of possible gains and losses for each player. We looked into two of such combinations. First, we investigated whether the second mover reacts to opportunities of joint improvements offered by the choice of the first mover. When a first mover chooses an opportunity set that allows for a Pareto improvement compared to the alternative opportunity set he could have chosen instead, the second mover might consider this as a proposal saying "let's make a deal". Our results indicate that whether the second mover reads such an intention from the first mover's choice or not, her behaviour is not affected by the presence of such mutually beneficial improvements as such. Second,

we investigated whether in situations typical to trust games, the choice of a trusting opportunity set by the first mover has a positive effect on the benevolence of the second mover. Here again, we find no evidence that the second mover's choice is affected in such situations beyond the effects of possible gains and losses. These two results are of interest to understand the specific motivations that shape second mover's behaviour in trust games.

Another significant contribution of our study is to show that incorporating a salient social norm, here the equality norm, can be useful to discriminate between different "revealed intentions". In its original formulation, the revealed intentions approach relies only on the comparison of actual choices to choices that would have been available but have not been made. However, shared social norms may create salient expectations which also affect behaviour. In such cases, it is simple to extend the revealed intentions approach by looking not only at how the opportunity set chosen by a player compares to the sets not chosen, but also at how it compares to the set of allocations suggested by the social norm. In our experiment, we find that this approach is useful to understand how some choices are interpreted by the second mover.

Overall, our study shows that it is possible to study a rich array of revealed intentions, without eliciting beliefs, by systematically varying the set of opportunity sets available to the first mover in a two-player two-stage game and by investigating the response of the second mover to the actual choice of the first mover and the alternative choice he could have made instead. This paper opens the path for further experimental work on revealed intentions. One may for instance consider that not only the possibility of gains and losses but their magnitude would have an influence on social preferences. Building on the present approach and

on the work by Fisman et al. (2007), further research might extend the findings presented here by investigating a richer set of revealed intentions using more complex choice sets than our, purposely simple, *lines* and *points*.

Chapter 6

Guilt-averse or reciprocal: Looking at behavioural motivations in the trust game

Abstract[1]

For the trust game, recent models of belief-dependent motivations make opposite predictions regarding the correlation between back-transfers and second-order beliefs of the trustee: While reciprocity models predict a negative correlation, guilt aversion models predict a positive one. This paper tests the hypothesis that the inconclusive results in previous studies investigating the reaction of trustees to their beliefs are due to the fact that reciprocity and guilt-aversion are behaviourally relevant for different subgroups and that their impact cancels out in the aggregate. We find little evidence in support of this hypothesis and conclude that type heterogeneity is unlikely to explain previous results.

[1]This study is co-authored by Rudolf Kerschbamer and Lionel Page.

6.1 Introduction

In social interactions, people frequently rely on others and take risks in the hope of reaching a more efficient outcome. Exchanges and trade typically involve an initial costly action by one party that benefits another person. On first sight, such an action seems kind and altruistic. On second thought, however, the underlying motivation driving such an act often stems from some expected benefit. The rule of reciprocation typically takes care of a beneficial outcome for both involved parties – even in one-shot situations. We frequently return favours, gifts and so like. In fact, not only in the English language the phrase "much obliged" has become a synonym for "thank you" indicating the feeling of indebtedness after receiving something. But what drives this feeling of obligation? Is it the other's expectation to receive something in return or is it the kindness of the act?

While documentation on reciprocal behaviour is vast, its motivation is less well studied and even less well understood. It is now well accepted that players' intentions and how these are perceived by the other players play a key role in explaining observed behaviour. To interpret the motivation behind an observed action, people have to form beliefs about the other's intention. As a consequence, it is difficult to derive and test theories concerning the basis of reciprocal pro-social behaviour purely from observed choices. Since the traditional game theory is not sufficient to describe many psychological or social aspects of motivation (Battigalli and Dufwenberg, 2009), theoretical explanations are typically based on the framework of psychological game theory that was developed by Geanakoplos et al. (1989). In this framework, preferences directly depend on beliefs about actions and beliefs: Players form higher-order

beliefs about each others' actions and their consequences on their own and on others' payoffs.

For second mover behaviour in the investment game, the two most prominent models of belief-dependent motivations make opposite predictions regarding the correlation between second-order beliefs and behaviour. According to the reciprocity theories of Rabin (1993) and Dufwenberg and Kirchsteiger (2004), a generous transfer by the first mover is interpreted by the second mover as less kind if the first mover is believed to expect a high back-transfer in return. These models therefore predict that the pro-sociality of the second mover *decreases* in his belief about the payoff expectation of the first mover. By contrast, the guilt aversion model introduced by Charness and Dufwenberg (2006) and generalized and extended by Battigalli and Dufwenberg (2007) assumes that people experience a feeling of guilt when they do not live up to others' (payoff) expectations. This model therefore predicts that the pro-sociality of the second mover *increases* in his second-order belief.

Given the conflicting predictions of the two classes of models, it is ultimately an empirical question whether high expectations (about the payoff expectation of the other) are detrimental or beneficial for pro-social behaviour. Previous studies investigating this issue – often obtained by employing variants of the trust game as the working horse – provide mixed results: While some papers (as for instance Guerra and Zizzo, 2004, Charness and Dufwenberg, 2006 and Bacharach et al., 2007) find a positive correlation between second-order beliefs and pro-social behaviour, many others (as for instance Strassmair, 2009, Ellingsen *et al.*, 2010, or Al-Ubaydli & Lee, 2012) find no correlation, or even a (slightly) negative one.

This paper explores the possibility that the inclusive evidence reported

in previous studies is due to preference heterogeneity in the population of second movers. Some subjects may be mainly motivated by reciprocity, some others by guilt aversion and a third group of subjects might not react to others' payoff expectations at all. If the former two groups are similar in size then in the aggregate the positive correlation between pro-social behaviour and second-order beliefs and the negative one might simply cancel out. This could explain the no-correlation result obtained in several previous studies.

To investigate this possibility, we use a triadic (that is, a three-games) design implemented within subjects. Our experimental design is intended to exogenously manipulate the second-order beliefs of trustees in the trust game and we use it to classify experimental trustees into behavioural types depending on how they react to the belief manipulation. In line with previous findings, we find no pronounced effect of the induced shift in second-order beliefs in the aggregate data. More importantly, we also do not find convincing evidence in support of our hypothesis that the no-correlation result in the aggregate data is caused by heterogeneity in second-mover preferences. Overall, it seems that the behaviour of second movers in the trust game is either not primarily driven by beliefs on the payoff expectations of the first mover or that it is driven by more complex considerations than those reflected in existing theories.[2]

The remainder of the paper unfolds as follows: Section 6.2 places our contribution in the existing literature. It is followed by the description of our experiment in Section 6.3 including our theoretical behavioural predictions. After defining several behavioural types in Section 6.4, we address our data and results in Section 6.5 before we conclude with a

[2]See Balafoutas et al. (2016) for evidence suggesting that more complex considerations than those implied by the model of simple guilt shape the pro-sociality of the "donor" in the standard dictator game.

discussion in Section 6.6.

6.2 Related literature

Social preferences have been modelled in a number of ways. As a first approach, outcome-based models have been suggested inter alia by Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Levine (1998). In these models, an agent's actions are solely motivated by the (distributional) properties of the outcome, i.e. how does the agent's own payoff compare to that of the other. The way in which a decision situation came about is irrelevant.

In models of intention-based social preferences, different intentions trigger variable responses to one and the same action. They thus take into account that motives influence the perception of another's action. Since preferences do not only depend on material payoffs but also on a player's interpretation of his opponent's behaviour, he has to form beliefs about the reason for a chosen action, i.e. the intention. In the literature of psychological game theory (Geanakoplos et al., 1989), players' utility depends directly on beliefs about others' choices or beliefs. Hereby, one does not only have to form beliefs about what the other person is going to do for the evaluation of the other's action but also about why he is going to it. Hence, one needs to form beliefs about what the other person believes oneself to do. The probably best known application of the psychological game theory is Rabin's (1993) theory of kindness-reciprocity: Players interpret other's behaviour belief-dependent as (un-)kind and act in turn (un-)kind themselves. His normal form model has been extended to extensive form games by Dufwenberg and Kirchsteiger (2004) and Falk and Fischbacher (2006). For the trust game, these kindness models

thereby predict (directly or indirectly) a negative relationship between one's second-order belief and one's reciprocity: Given a certain distributional outcome, the higher one's second-order belief (i.e. the more one beliefs the other hopes/expects to accrue for himself from the final payoff allocation), the less kind the other's choice is interpreted and in turn the less kind one's own behaviour.[3] Another suggested belief-dependent preference is guilt aversion. According to Baumeister et al. (1994), people feeling guilt for failing up to their partner's expectations will alter their behaviour (to avoid guilt) in ways that seem likely to maintain and strengthen their relationship. Building on this statement, associated models of guilt aversion (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007; Vanberg, 2008) assume that an agent experiences some disutility (guilt) if he does not live up to the other's expectations which in turn affects his utility influencing his decision-making.[4] Hence, given a certain choice, the higher one's second-order belief, the more generous one's action in order to avoid the feeling of guilt.

While the reciprocity based on kindness theory as well as the guilt aversion theory are in general able to explain reciprocal behaviour that is observed in many experiments and specifically in the trust game, they are distinct in the way that second-order beliefs affect choices. In fact, predictions go in opposite directions. However, both theories are difficult to test as the manipulation just as the elicitation of beliefs is challenging.

[3]Note that the model by Rabin (1993) does not explicitly incorporate second-order beliefs but one's kindness is a function of the belief about what the other wants oneself to receive. In a zero-sum game (such as the second stage of the trust game), this notion is easily transferred into the notion of second-order beliefs: If a player believes that the other wants oneself to receive less, this is equivalent to believing that the other wants to receive more himself. In this way, also Rabin's model on reciprocity predicts a negative relationship between second-order beliefs and own kindness.

[4]A person however only feels obligated to fulfil payoff/outcome expectations rather than action expectations.

Yet, they both strongly rely on beliefs about other's payoff expectations. Therefore, it may not be surprising that the experimental evidence is not conclusive.

Stanca et al. (2009) compare reactions to intrinsically versus extrinsically motivated actions. In particular, they investigate how second movers' back-transfers in a gift-exchange game relate to their back-transfers in a slightly modified game, in which first movers did not know about second movers' possibility to send money back when making their decision. They find an increased positive reciprocity when first movers' transfer was exclusively driven by intrinsic motivation (no information treatment). However, on average, there is no increase in the back-transfer. It is rather the slope of second movers' average reaction function that is steeper. Stanca (2009) makes a similar comparison between a treatment of direct reciprocity (A helps B then B helps A) and a treatment of generalized indirect reciprocity (A helps B then B helps C) in which transfers made by the first mover are solely intrinsically motivated. He finds that B's transfer is on average significantly larger and his reciprocal behaviour is stronger pronounced (i.e. steeper slope in second mover's average reaction function) in the latter. Overall, his results fit the kindness-based reciprocity theory quite well.

Testing directly for guilt aversion, existing studies typically measure the correlation between players' second-order beliefs, their belief about what others expect them to do, and players' actions. The second-order expectations are usually obtained by their direct elicitation from second movers. Studies with such a belief manipulation often report (strong) support for the guilt aversion hypothesis (Charness and Dufwenberg, 2006; Dufwenberg and Gneezy, 2000; Bacharach et al., 2007; Guerra and Zizzo, 2004). This method might however entail the risk of an endogene-

ity bias because the explanatory variable is not randomly manipulated. A possible caveat is the (false) consensus effect (Ross et al., 1977). This effect occurs if one believes that others behave similar to oneself and thus that one also believes that others expect a behaviour that is similar. If one makes a large transfer, one also thinks that others expect this. Players' second-order beliefs are then influenced by their actions rather than vice versa. Bellemare et al. (2011) test the importance of the (false) consensus effect and indeed find that the estimation of the willingness to pay to avoid guilt can be substantially overestimated if stated belief data is being used and the correlation between stated beliefs and preferences is not accounted for. Nevertheless, the willingness to pay remains significantly positive once using an exogenous belief manipulation, which supports the guilt aversion theory. Similarly, Ellingsen et al. (2010) had already tried to establish causation by revealing actual first-order expectations to second movers.[5] However, they find no significant effect of expectations on the paired player's reciprocation – neither in a dictator nor in two versions of the trust game. Instead, they conclude that the consensus effect is responsible for the major fraction of the correlation between second-order beliefs and behaviour previously found and that guilt aversion is smaller than thought. Another study manipulating beliefs exogenously was conducted by Al-Ubaydli and Lee (2012). In their first treatment, they disclose the expectations of neutral observers to second movers and use the guesses as instruments. In the second treatment, they adopt the design used by Ellingsen et al. (2010) and transmit expectations to each first mover's partner. They,

[5]The authors elicited these first-order beliefs in a way assuring maximized honesty. First movers were asked to guess the outcome of the game (incentivized) while not telling them that their guesses would be revealed to second movers. Second movers were informed about the beliefs and knew about the nescience of their transmission.

again, find no effect of expectations on reward or punishment behaviour in either treatment. Only when they elicit expectations directly, they find a correlation supporting guilt aversion. Closest to our paper, is the study by Strassmair (2009). In her modified trust game, second movers can only reciprocate with some exogenous probability. Varying the probability allows her to manipulate expectations: if it is high, the first mover can expect a back-transfer more often than if the probability is low. Hence, his behaviour is more likely to be "selfishly" motivated and less kind. Again, her between-subject design reveals no indication that expectations influence behaviour in either way. Overall, the experimental evidence raises doubt on the early findings of guilt aversion as the found correlation between expectations and behaviour seems to be mainly driven by an endogeneity bias and thus vanishes once beliefs are manipulated exogenously. However, evidence in support of the theories of reciprocity based on kindness is also rare.

In line with the zero correlation findings, Vanberg (2008) reports that while people are prone to keep their own promises, they are reluctant to fulfil expectations if the promise was made by someone else. This lead Ederer and Stremitzer (2014) to the assumption that an agent might only be affected by guilt aversion if he himself is responsible for causing the expectations for example by making a promise. Their experimental design resembles Strassmair's (2009) and our's but includes a communication stage. They do find support for their claim: Expectations (and guilt) only mattered if they were supported by a promisory link between the two acting parties. While this is an interesting explanation of the often found zero correlation, the findings of Kawagoe and Narita (2014) raise concerns about its validity. They test exactly for this "personal

guilt aversion"[6] using a trust game with hidden action and transmitting beliefs from the first mover to the second mover following the design by Ellingsen et al. (2010) and Reuben et al. (2009). Similar to Ederer and Stremitzer (2014), they have treatments with and without pre-play communication. In contrast to the prediction by the guilt aversion theory as well as the personal guilt aversion hypothesis, elicited beliefs do not (positively) correlate with reciprocal behaviour. In particular, even in the communication treatment, no such correlation is found. In fact, they report a slight positive correlation in the no communication treatment although it is not significant.

Based on the mixed evidence, we think it is worthwhile considering another potential cause of the "missing" impact of expectations on choices: the existence of several types (kindness-reciprocal and guilt-averse) whose behavioural differences cancel out on average. Psychological motivations thus remain undiscovered in between-subject designs. We base this hypothesis on several findings in previous studies which were typically only noted as side remarks. Reuben et al. (2009) for instance investigate second movers' behaviour in a lost wallet game. They compare their back-transfers when they do not have any information about the partnered first mover's expectation with their behaviour after observing a low/high expectation. While they observe a significant decrease in the average back-transfer after observing a low expectation, the increase following an observation of a high expectation remains insignificant. They interpret their data as evidence for guilt aversion but also report quite diverse and distinct reaction patterns. In both treatments, more than 10 percent behave in a way consistent with the kindness-reciprocity theory by in-

[6]Personal guilt aversion postulates that people feel guilty when they betray another person's expectation, with that expectation having been raised by their very own actions (typically by their promises).

creasing their back-transfer when confronted with a low expectation and vice versa. In addition, a substantial fraction of players does not react at all to expectations. Furthermore, Bellemare et al. (2015) test for heterogeneities in a (mini) dictator game. Although, their work is focused on and limited to guilt sensitivities, they provide a further indication of diverse personal differences in reaction patterns to second-order beliefs. The study by Attanasi et al. (2013) pursuits a similar goal to us. They determine belief-dependent preferences from an unincentivized questionnaire and confirm the presence of different types: more than 50 percent behave consistent with guilt aversion, 16 percent with the reciprocity theory based on kindness, 6 percent show a mixed reaction pattern and the rest does not react to a change in the other's expectations (15 percent are motivated by self-interest only and 9 percent always return a hypothetical positive but constant amount).

6.3 The experiment

6.3.1 Experimental design

The game

We employ a triadic (three-games) design implemented within subjects to manipulate the second-order beliefs of experimental trustees in a binary investment game. The structure of each of the three games is as illustrated in Figure 6.1.[7] There are two players – a first mover (FM, he) and a second mover (SM, she). The players start with identical initial endowments of \$10 (all amounts are in Australian dollars). In the first

[7] A similar experimental design has previously been employed by Strassmair (2009) in an across-subjects study.

stage, the FM decides between keeping the endowment and sending the amount of \$3 to the SM. If the FM decides to keep the endowment, the game ends and both players receive their endowments of \$10 as their final payoffs. If the FM transfers the amount of \$3, this amount is multiplied by 5 and the resulting \$15 are then credited to the account of the SM. Now, a random move by Nature determines whether the game stops. With the probability $1 - p$, the state of the world is $\omega = 0$ and the game stops. In this case, the FM receives the \$7 that are left from his initial endowment and the SM receives her initial endowment plus the \$15 from the transfer of the FM. With probability p, the state is $\omega = 1$ and the game continues. In this case, the SM can now decide how much money she wants to send back to the FM. She can choose any integer amount x between 0 and 15. The FM then receives the \$7 that are left from his initial endowment plus the SM's back-transfer x as the final payoff. The SM earns her initial endowment (\$10) plus the multiplied transfer (\$15) minus the amount x she has chosen to send back to the FM. At the end of the game, both players learn their payoffs and the outcome of Nature's move (i.e. whether the game was stopped or the SM had the opportunity to make a back-transfer).

The crux of our working horse trust game consists in the random move by Nature after the FM's sending decision. The game resembles a standard binary trust game if $p = 1$, as the SM can then make a back-transfer with certainty. By contrast, for $p = 0$, the game is reduced to a dictator game (with the FM as the dictator). To manipulate the belief of the SM about the payoff expectation of the FM (conditional on sending the amount of \$3), we vary – across treatments – the probability p that the SM can make a back-transfer, while keeping everything else constant.[8]

[8]Note that by keeping the FM's transfer constant across the three treatments, we

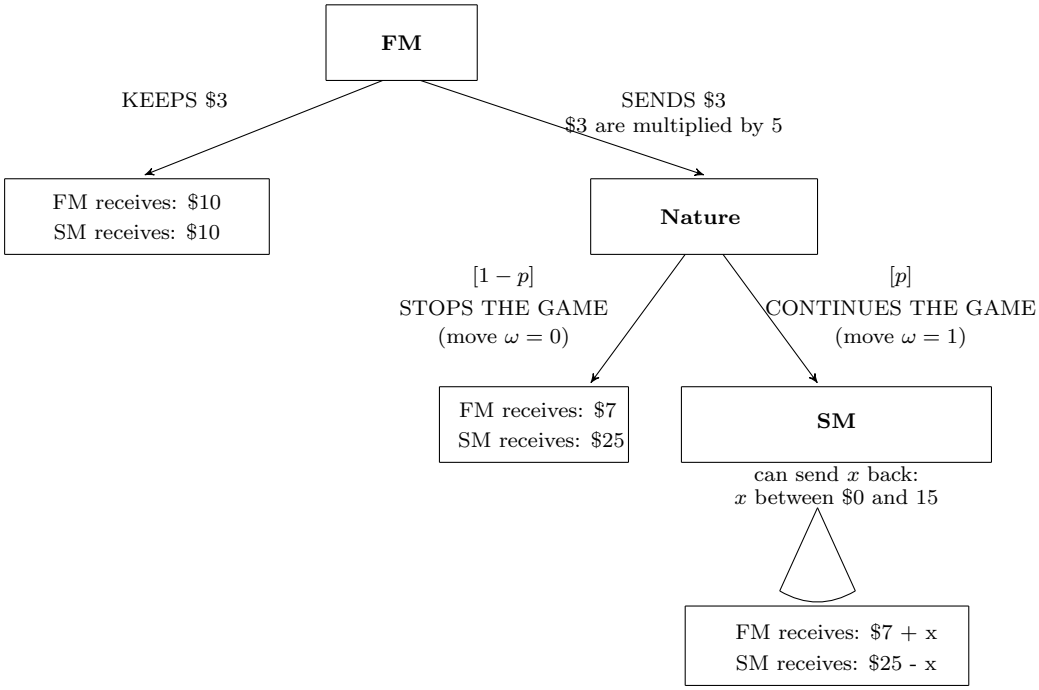


Figure 6.1: Structure of our modified trust game.

The manipulation of the continuation probability p essentially allows an *exogenous belief manipulation*: the lower p, the lower the chance that the FM will receive some money back from the SM, the lower therefore arguably his payoff expectation conditional on making the transfer of \$3, the lower therefore also the expectation of the SM on the payoff expectation of the FM. Conversely, the closer p is to 1, the higher the chance of a back-transfer from the SM, the higher therefore arguably also the SM's belief about the payoff expectation of the FM. Because we are interested in individual response patterns, every subject has to make a choice in three treatments differing only in the continuation probability p. A subject in the role of the FM is asked whether he wants to make the transfer of \$3 in each treatment. According to the game tree in

keep the SM's choice set constant across treatments. This seems important to control for distributional concerns that might shape the back-transfer of the SM.

Figure 6.1, whether or not the SM has a decision to make depends on the FM's choice and on Nature's random move. To collect data from all subjects in all treatments, we apply the strategy method. That is, subjects in the role of the SM are asked to make a decision regarding the back-transfer assuming the FM made the transfer and Nature did not stop the game. To make the SM's decision scenario plausible in each of the three treatments we decided to make the choice of the initial transfer by the FM quite attractive by using high values of p. Furthermore, we wanted to avoid degenerate lotteries while making the difference between the decision tasks as distinct as possible. Specifically, the variable p takes on the values 50, 70 and 90 percent across our three treatments.

The Observer

The experimental design is intended to manipulate the belief of the SM about the payoff expectation of the FM (conditional on sending the amount of \$3). To verify that this manipulation works (i.e. that a higher continuation probability is associated with higher payoff expectations of the FM), we have a third player role in our experiment, the role of an impartial observer. The task of the Observer is to guess how much money the participants in the role of the SM send back, on average, to the paired FM assuming that the FM transferred the \$3 and Nature did not stop the game. From these joint conditional beliefs, we can then calculate the expectation of the Observer about the expected payoff associated with the initial transfer by the FM for each of the three treatments. We can then check if and how this expected payoff varies with the continuation probability. We use an impartial observer to elicit beliefs to avoid the usual problems associated with eliciting beliefs from

agents that also have to make a decision.[9]

6.3.2 Experimental procedure

The experiment was conducted between February and June 2015. To the 15 experimental sessions, we recruited 180 students from a large university in Australia via the ORSEE software (Greiner, 2015). Each session lasted approximately 45 minutes. No participation fee was paid and the average earnings were $14.30. The experiment was programmed and conducted with the experimental software CORAL (Schaffner, 2013). At the beginning of the experiment, each participant was randomly assigned the role of either the FM, or the SM or the Observer and participants kept the role during the entire session. After session 10, we disposed the role of the Observer because we attained enough data to test whether our belief manipulation worked. At no time were subjects informed about the identity of their matched partner. The full instructions can be found in Appendix C.

In each session, participants where exposed successively to the three treatments distinguished only in the continuation probability p. Subjects received neither any feedback on the choices made by other participants nor on the outcome of Nature's move before all decisions were made. At the end of the experiment, one of the three treatments was randomly

[9]If beliefs are elicited before the decision is made, this might lead to an "experimenter demand effect", or to a "consistency effect": Subjects might condition their choice on the stated belief because they believe that the experimenter expects them to do so, or actions might be shaped by beliefs just to be consistent. Fleming and Zizzo (2015) test the impact of the experimenter demand effect on choices in a different context and indeed find convincing evidence in line with it. By contrast, if beliefs are elicited after the choices, then actions might influence (or cause) beliefs. This is often referred to as the "projection hypothesis", or the "false consensus effect". Bellemare et al. (2011) test the importance of the (false) consensus effect and indeed find evidence in line with it.

selected for payment. The players' actions as well as the move by Nature for that particular treatment were revealed and payoffs calculated accordingly.[10] The beliefs of subjects in the role of the Observer were incentivized using the quadratic scoring rule. Specifically, we implemented the payoff function

$$\text{Payoff}_{\text{Observer}} = 15 - 0.5(\bar{x} - x_{\text{Guess}})^2,$$

where $\bar{x}$ is the rounded average back-transfer made by subjects in the role of the SM and x_{Guess} is the Observer's associated guess.

6.3.3 Theoretical predictions

In our experimental design and the trust game in general, a positive transfer by the first mover always seems altruistic/kind on first sight: the act is costly (deduction from his endowment) and benefits the second mover. While this is true for all continuation probabilities, the FM's chance to receive a future return for his transfer changes because the SM's possibility of a back-transfer is varied. Thus Hypothesis 6.1:

Hypothesis 6.1

Given an initial transfer, the FM can expect a higher return the higher the continuation probability, i.e. $E(x_1|p = 50\%) < E(x_1|p = 70\%) < E(x_1|p = 90\%)$.

Obviously, the SM could interpret a transfer kinder if p is low and thus adjust her back-transfer. Consequently, such a more generous behaviour generally could offset the smaller p at least partially. Yet, it should not

[10] The SM's decision was only revealed to the FM if the FM sent the \$3 and Nature did not stop the game.

lead to opposing predictions. To see this, let's assume the opposite for a moment: If the FM expects ex-ante higher rewards for a small continuation probability, then the transfer should be interpreted as less kind for small p and rewarded less generously. Thus, the FM's expectations were incorrect. Nevertheless, we provide more evidence on the relationship between the continuation probability p and the FM's ex-ante expectations by eliciting beliefs from impartial observers (see Section 6.3.1).

In what follows, we discuss the most commonly used models of (social) preferences. They all differ in their underlying assumptions and the role of expectations on an agent's utility function and thus differ in their behavioural predictions in the investigated modified trust game.

Neoclassical theory

In the standard neoclassical theory, beliefs about other's behaviour play a role only in so far as, in equilibrium, they should be correct. Belief-dependent motivational interpretations, contrariwise, are neglected and do not influence behavioural predictions. Instead, subjects are assumed to solely maximize own monetary income and to act rationally. Given these assumptions, the SM's decision will not depend on the continuation probability p. Rather, in the unique subgame-perfect Nash equilibrium, she would always choose to keep the whole surplus for herself.

Outcome-based social preferences

Acknowledging the fact that people may not only be own-income maximizing but also be motivated by fairness considerations, models of outcome-based social preferences assume that the distribution of all players' income influences an agent's utility function. In our experimental design, the FM's transfer is kept constant across all continuation probabilities

p. Since the SM's choice is influenced by outcomes only, she faces the same decision problem at her decision node independent of p. Her back-transfer may therefore take positive values, yet will not vary across decision tasks.[11]

Simple guilt aversion

Guilt in the context of social preferences refers to a guilt feeling as the consequence of not living up to others' expectations (Baumeister et al., 1994; Battigalli and Dufwenberg, 2007). In what follows, we focus on the theory of "simple guilt" proposed by Charness and Dufwenberg (2006) and extended by Battigalli and Dufwenberg (2007). The guilt aversion hypothesis postulates that players experience a utility loss if they believe that they let other's payoff expectation down.[12] The basic idea is that the SM suffers from guilt to the extent that she believes that the FM gets a lower monetary payoff than the SM believes the FM expects to receive. The higher the SM's second-order belief, her belief about the FM's payoff expectation, the larger her experienced disutility from a certain payoff allocation assigning the FM less money than he is believed to expect. Consequently, the SM will adjust her choice in order to mitigate or even avoid the experienced guilt. In our design, a guilt-averse SM is therefore predicted to increase her back-transfer as the continuation probability p rises.

[11] An individual exhibiting outcome-based social preferences is not necessarily altruistic or inequality-averse. Also spiteful, envious or inequality-seeking agents integrate the other agent's payoff in their maximization problem. However, in our design, these types would behave exactly as an agent only motivated by self-interest.

[12] In contrast, the theory of "guilt from blame" (Battigalli and Dufwenberg, 2007, 2009) postulates that players experience a utility loss if they believe that *others believe that they believe* that they let others' payoff expectation down.

Kindness-reciprocity

Models of intention-based social preferences typically focus on an agent's reciprocity based on his kindness interpretation (e.g. Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006). The perceived kindness of a certain action thereby depends on the interpretation of the other's intention. The kinder the perceived intention, the kinder the agent's subsequent decision. In zero-sum games, second-order beliefs (what the SM believes the FM expects to be the final outcome) shape an agent's intention perception: The more the SM thinks the FM hopes to accrue for himself, the less kind the FM's action is perceived. So, if the FM makes the transfer because he is hoping to gain himself substantially from this move, i.e. he has high payoff expectations, the SM's back-transfer should be lower than if she beliefs the FM's transfer to be motivated by true kindness, i.e. he has low payoff expectations. According to Hypothesis 6.1, a higher continuation probability p will, ceteris paribus, increase the FM's payoff expectations. The action of a transfer should therefore be interpreted as kinder if p is low and as less kind if p is high. A SM motivated by such reciprocal preferences will adjust her back-transfer accordingly: A SM with reciprocal preferences decreases her back-transfer as the continuation probability rises.

6.4 Behavioural types

To describe and distinguish individual behavioural patterns, we define four types of players – selfish (S), altruistic (A), guilt-averse (G) and reciprocal (R) ones. For each of these types, we assume a linear relationship between the continuation probability and the back-transfer. Specifically,

the back-transfer of a SM of type $t \in \{S, A, G, R\}$ is assumed to be a function of her unconditional altruism parameter c_t and of a parameter m_t which reflects how she reacts to our belief manipulation:

$$x_t(p) = c_t + m_t p \tag{6.1}$$

Definition 6.1 (Selfish Agent) *A SM is said to act in a selfish manner if her back-transfer is always zero:* $c_S = 0$ *and* $m_S = 0$, *implying* $x_S(p) = 0$ *for all* p.

Definition 6.2 (Unconditional Altruist) *A SM is said to be an unconditional altruist if her choice is unaffected by her belief about the payoff expectation of the FM but she nevertheless returns a positive amount. Thus, her back-transfer* x *is a constant amount independent of the continuation probability* p: $c_A > 0$ *and* $m_A = 0$, *implying* $x_A(p) = c_A$ *for all* p.

Definition 6.3 (Guilt-Averse Agent) *A SM is said to be guilt-averse if her pro-sociality is increasing in her belief about the payoff expectation of the FM. Thus, her back-transfer* x *is an increasing function of the continuation probability* p: $c_G \geq 0$ *and* $m_G > 0$, *implying* $x_G(p) = c_G + m_G p$ – *with* $m_G > 0$ – *for all* p.

Definition 6.4 (Reciprocal Agent) *A SM is said to be reciprocal if her pro-sociality is decreasing in her belief about the payoff expectation of the FM. Thus, her back-transfer* x *is a decreasing function of the continuation probability* p: $c_R \geq 0$ *and* $m_R < 0$, *implying* $x_R(p) = c_R + m_R p$ – *with* $m_R < 0$ – *for all* p.

6.5 Data and results

In total, we collected data from 180 students – 70 subjects in the role of the FM, 70 subjects in the role of the SM, and 40 subjects in the role of the Observer. Since each subject made a decision in each of the three treatments, we have 210 observations for the role of the FM, 210 observations for the role of the SM, and 120 observations for the role of the Observer.

6.5.1 The Observer

To confirm the validity of our experimental belief manipulation, we first look at the data obtained from subjects in the role of the Observer. We first investigate their guesses about the average back-transfer and compare guesses with actual behaviour. As can be seen from Figure 6.2, the Observers' average guesses are roughly \$1 higher than the actual choices of SMs for all continuation probabilities. However, this difference is not significant for any of the treatments (the Mann-Whitney ranksum test[13] p-values are 0.0596, 0.1639 and 0.1619 for the continuation probabilities of 50%, 70% and 90%, respectively) so that the Observers' guesses are on average a decent approximation of actual behaviour.

Further, we can see a slight upwards trend in guesses as the continuation probability increases. Yet, the differences in average beliefs across the three continuation probabilities are not statistically significant (the Wilcoxon signed-rank test p-values are 0.3448 for H_0: $E(x|p = 50\%) = E(x|p = 70\%)$, 0.3180 for H_0: $E(x|p = 70\%) = E(x|p = 90\%)$ and 0.2468 for H_0: $E(x|p = 50\%) = E(x|p = 90\%)$).

[13]Comparisons between the different player groups are unmatched so that we use the two-tailed Mann-Whitney ranksum test.

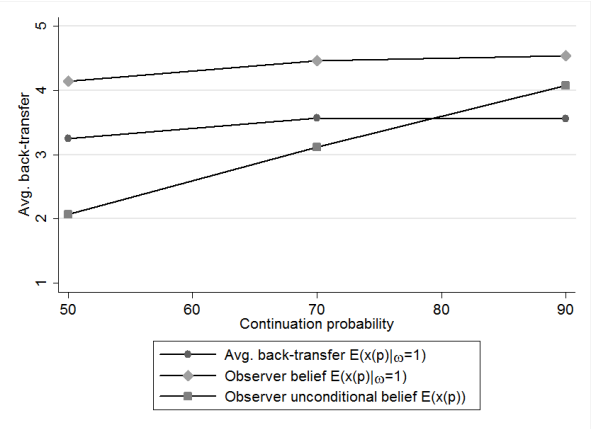


Figure 6.2: Average back-transfers by the SMs as a function of the continuation probability p compared to the Observers' average guess and the associated expected return for the FM conditional on making the transfer.

It is important to note that we have elicited joint conditional beliefs about average back-transfers. Specifically, subjects in the role of the Observer were asked how much they thought the SM would on average transfer back, assuming the FM transferred the \$3 and Nature did not stop the game. We are however interested in preferences which are influenced by the (belief of the SM on the) payoff expectation of the FM conditional only on the own decision (of sending the \$3). To obtain information on this expectation, we multiply the joint conditional belief by the continuation probability p. The resulting number $\widetilde{x}_1^O$, estimated from Observers' guesses, is significantly increasing in p: $E(\widetilde{x}_1^O|p = 50\%) = 1.86 < E(\widetilde{x}_1^O|p = 70\%) = 2.78 < E(\widetilde{x}_1^O|p = 90\%) = 3.67$ (Wilcoxon signed-rank test, p-values < 0.01). Assuming that Observers' beliefs are a good approximation of real players' beliefs, we interpret this result as evidence showing that our belief manipulation works.

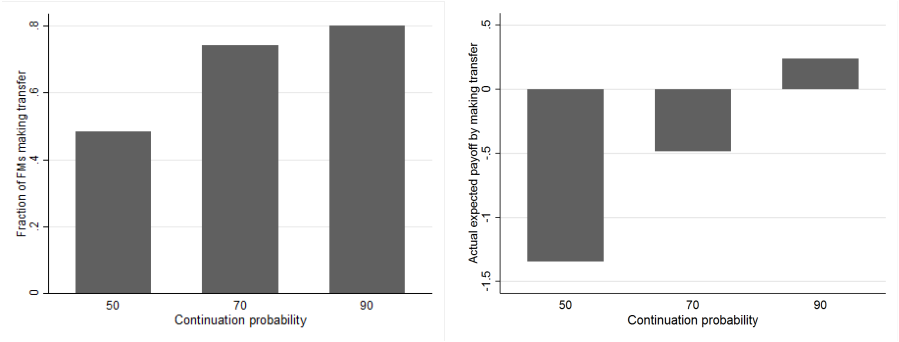


Figure 6.3: Left panel: Fraction of FMs making the transfer for each of the three continuation probabilities. Right panel: FMs' average payoff conditional on making the transfer for each of the three continuation probabilities.

6.5.2 The first mover

We now turn to the data obtained from experimental FMs. The left panel of Figure 6.3 shows the fraction of FMs making the transfer for each of the three continuation probabilities. Over 50 percent make the transfer independent of p, but there is a clear increase in the fraction as p increases – more FMs send the money when the probability that the SM can actually send a back-transfer is higher. This is a further indication in support of our main hypothesis that the payoff expectation of the FM (conditional on sending the \$3) is increasing in p. As can be seen from the right panel of Figure 6.3, making the transfer pays off, on average, only when the continuation probability is 90%. This reveals that even if SMs' back-transfers were on average higher for a low p, they do not offset the higher risk taken by the FM. FMs' payoff expectations based on actual back-transfers should therefore be increasing with the continuation probability which is in line with our main assumption stated in Hypothesis 6.1.

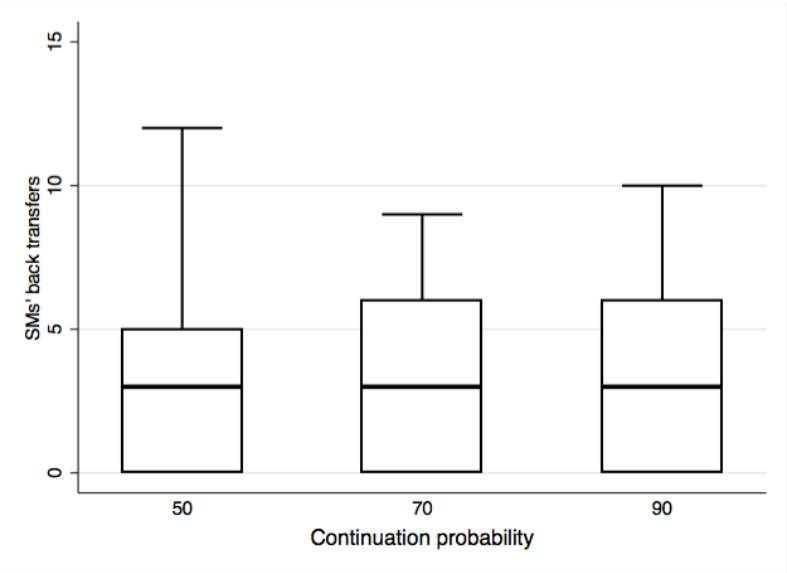


Figure 6.4: Boxplots of SMs' back-transfers depending on the continuation probability p.

6.5.3 The second mover

We now turn to our main data source, the data obtained from experimental SMs. First, we look at average back-transfers. Figure 6.2 shows that average SM behaviour is quite similar across the three continuation probabilities. Statistical tests confirm that average back-transfers are not significantly different across treatments (the Wilcoxon signed-rank test p-values are 0.0822 for H_0: $E(x|p = 50\%) = E(x|p = 70\%)$, 0.3518 for H_0: $E(x|p = 70\%) = E(x|p = 90\%)$ and 0.0451 for H_0: $E(x|p = 50\%) = E(x|p = 90\%)$). Similarly, the distributions of choices do not vary across p (Kolmogorov-Smirnov test, combined p-values: 0.959 for H_0: $\Phi(x|p = 50\%) = \Phi(x|p = 70\%)$, 0.959 for H_0: $\Phi(x|p = 70\%) = \Phi(x|p = 90\%)$ and 0.751 for H_0: $\Phi(x|p = 50\%) = \Phi(x|p = 90\%)$). The respective boxplots are shown in Figure 6.4. These results are in line with the no-correlation results obtained in several previous studies (cf. Section 6.2).

Looking at individual behaviour, it can be noted that SMs' choices appear quite heterogeneous so that we cannot determine one clear pattern. The associated individual graphs are displayed in Appendix C. The heterogeneity in reactions to changing continuation probabilities is in line with our research hypothesis of diverse types: Some participants do not respond to changing continuation probabilities (and thus different expectations) at all, while others behave consistent with the guilt aversion theory, and again others seem to follow the predictions of the model of kindness-reciprocity. However, quite a substantial fraction of subjects cannot be easily classified (around 26 percent).

To get more insights into individual behaviour, we run a mixture model (Harrison and Rutström, 2009), which allows us to estimate the fraction of subjects whose choices are consistent with one of the types defined in Section 6.4. The mixture model allows different types to coexist in the same sample and it determines the support for each of the types indicating their respective importance in the population. To simplify the estimation procedure of the mixture model, we decided to identify and exclude the selfish agents manually as they can easily be determined. We ended up removing 15 individuals from our dataset who never returned any money (participant numbers: 8, 29, 32, 35, 47, 83, 95, 110, 113, 119, 126, 132, 158, 164, 56 in Appendix C), and four agents who returned \$1 once and zero otherwise (participant numbers: 17, 26, 128, 138). Hence, 27 percent do not exhibit any kind of social preferences.[14] Using the definitions in Section 6.4, we specify one likelihood function for the remaining competing types $t \in \{A, G, R\}$, conditional on the respective model being correct:

[14]We also run the mixture model including the selfish types where they would form a "neutral" type together with the unconditional altruists. The higher likelihood was however reached by excluding them.

$$\ln L_t(x, c_t, m_t, \sigma) = \sum_i \ln l_{ti} = \sum_i \ln[\Phi_t(x_i)],$$

where m_t is restricted: $m_A = 0$, $m_G > 0$ and $m_R < 0$. Our grand likelihood of the entire model is then the probability weighted average of the conditional likelihoods, where π_t denotes the probability that the respective type applies and where l_{ti} is the respective conditional likelihood:

$$\ln L(x, c_t, m_t, \sigma, \pi_t) = \sum_i \ln[(\pi_A \times l_{Ai}) + (\pi_G \times l_{Gi}) + (\pi_R \times l_{Ri})].$$

The parameter estimates can directly be found by maximizing this log-likelihood. Table 6.1 presents the resulting maximum-likelihood estimates of the mixture model.[15] The first finding is that the estimates for the probabilities of our type specifications are all positive and significantly different from zero. Their respective size refers to the fraction of choices characterized by each. For the data at hand, guilt aversion seems to dominate slightly – with 46 percent – but closer inspection reveals that we cannot reject the hypothesis that the three probabilities are identical (p-values: 0.1100 for H_0: $\pi_A = \pi_G$, 0.0815 for H_0: $\pi_G = \pi_R$ and 0.9359 for H_0: $\pi_A = \pi_R$). Yet, looking at the estimation results reveals very flat slopes for both, reciprocal ($m_R = 0.007$) and guilt-averse types ($m_G = -0.024$). Figure 6.5 graphically illustrates these findings. It shows – for each of the three types – the plot of the estimated function of the back-transfer on the continuation probability. Although, there seem to be behavioural tendencies present, the effect of a change in the continuation probability seems to be rather weak, especially for guilt-averse

[15] Our likelihood function seems to have several local maxima. The results reported here refer to the estimates with the highest likelihood found.

agents. But also the effect for reciprocal agents is not very pronounced.

Mixture Model (N=153): $\ln L(x, c_t, m_t, \sigma, \pi_t) = \sum_i \sum_t \ln[(\pi_t \times l_{ti})]$

Parameter	Estimate	Robust SE
c_G	3.008***	.742
c_R	8.881***	.955
c_A	1.236**	.424
m_G	.007**	
m_R	-.024***	
σ	1.161	
π_G	.464***	.069
π_R	.273***	.062
π_A	.293***	.071

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6.1: Maximum-likelihood estimates of mixture model.

Since we do not interpret these results as convincing evidence in support of our hypothesis of the coexistence of guilt-averse and reciprocal agents, we next try another approach to test for the presence of heterogeneity in the reaction to the second-order belief. Specifically, we estimate two versions of a linear regression model of the back-transfer on the continuation probability. One model allows only for random intercepts while the other allows for random intercepts *and* random slopes. Our "random-intercept" model reads:

$$x_i(p) = c + \beta p + u_{0i} + \epsilon_i,$$

where x_i is subject i's back-transfer, c is a constant, p is the continuation probability and u_{0i} is the subject-specific random effect. The "random-slope" model – allowing the intercept *and* the slope to vary between participants – reads

$$x_i(p) = c + \beta p + u_{0i} + u_{1i} p + \varepsilon_i,$$

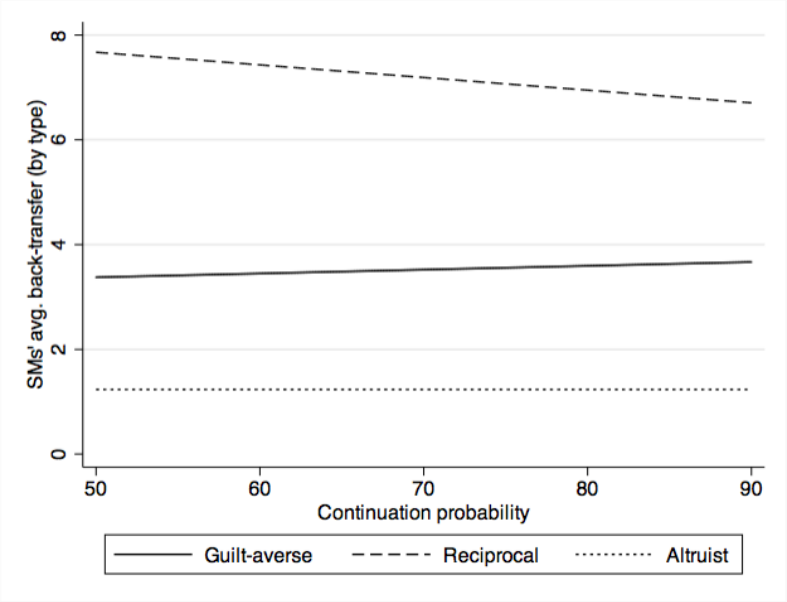


Figure 6.5: Plot of the estimated type-functions based on the estimates of the mixture model.

where u_{1i} is the additional subject-specific random effect on the slope of p. The results for both models are reported in Table 6.2. The estimates of the "fixed" parameters confirm the results obtained from the mixture model: The constant c is positive and significant but the effect of p on back-transfers is insignificant. Our main interest lies in the results obtained for σ_{u_0} and σ_{u_1} as they represent the between-subject variation in the intercept and the slope of p, respectively. The significance of σ_{u_0} can be tested using the likelihood ratio (LR) test of the linear regression model in its restricted version of the random-intercept model. The null hypothesis that $\sigma^2_{u_0}$ is zero can be rejected at the 0.01 percent significance level (p-value < 0.0001). To test the significance of σ_{u_1}, we again use a LR test. This time, we test the random-slope model against the random-intercept model. The p-value is 0.2116 so that we cannot reject the null hypothesis that $\sigma^2_{u_1} = 0$ and thus that the slope of the back-transfer as a function of the continuation probability p is the same for all subjects.

Multi-level Models (N=210): $x_i(p) = c + \beta p + u_{0i} + u_{1i}p + \varepsilon_i$

	Random-intercept model		**Random-slope model**	
Parameter	Estimate	Robust SD	Estimate	Robust SD
p	.007	.007	.007	.007
c	2.988***	.609	2.988***	.578
Random effects				
σ_{u_1}			.018	.008
σ_{u_0}	2.746***	.262	2.456***	.359

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6.2: Mixed-effects maximum-likelihood estimates of multi-level models.

6.6 Discussion

We have experimentally investigated the empirical relevance of the most prominent models of belief-dependent motivations for behaviour in the binary trust game. We tested how subjects respond to another player's payoff expectations given a certain distributional outcome. Based on mixed results in previous studies, we hypothesized that the behaviour observed in the different versions of the trust game is generated by heterogeneous mental processes: guilt aversion, kindness-reciprocity, unconditional altruism and own-payoff maximization. In particular, we argued that the effects of the first two are confounded and even masked when looking at averages as their associated behavioural predictions go in opposite directions. Our triadic design implemented within subjects has allowed us to study individual response patterns to exogenously manipulated second-order beliefs. Our results however indicate that individual differences in reactions to the other agent's payoff expectations may not be as important as suggested by our research hypothesis. Results obtained from a mixture model allowing for reciprocal and guilt-averse agents as well as for unconditional altruists suggested that individual differences exist only in the *level* of exhibited pro-social behaviour. The effect of the induced

change in second-order beliefs on choices was found to be negligible – on average *and* on the type level. We have confirmed these findings by estimating two versions of a random-coefficient model allowing the reaction of the SM to the belief manipulation to differ within our sample. While we found support for heterogeneity in the level of unconditional altruism, we do not find convincing evidence for heterogeneity in how second movers react to the induced shift in their second-order beliefs. Our results suggest that the most prominent models of belief-dependent motivations – reciprocity and aversion against simple guilt – may not accurately reflect how players in the role of the second mover in the trust game react to their beliefs about the payoff expectation of the first mover. Further work is needed in this area to understand the role played by higher-order beliefs for behaviour.

Part III

Conclusion and Discussion

Chapter 7

Summary and Possible Extensions

This thesis investigated the role of intentions on people's decision-making applying conceptual and methodological insights from behavioural and experimental economics. Together, the three studies in this dissertation enrich the existing knowledge on intentions and the motivational channels through which they become effective. Each of the presented essays was devoted to reciprocity in its broadest sense but investigated it in different settings and analysed it from slightly different angles. Each study thereby focused on different channels through which intention-based social behaviour may be triggered. This thesis looked at various intentions determining perceived kindness and investigated their judgement as well as their effect on another's behaviour through reciprocity and trustworthiness.

The study *Driving a hard bargain is a balancing act: The importance of reciprocity in bargaining* (Chapter 4) investigated the role of

reciprocity in a simple bargaining setting. As bargaining is an ubiquitous process in all our lives, it is of great interest which strategy helps to achieve the best possible outcome for oneself and to avoid an impasse. Our results suggest that it is not advisable to only focus on one's own outcome and that it is also not sufficient to look out for the other's outcome. Instead, it is crucial to also pay attention to the bargaining *strategy* one adopts. We argued that players do not enter a bargaining situation with a predetermined minimal acceptable payoff outcome, but that the bargaining process itself influences an agent's valuation of what is acceptable. Specifically, offers and counter-offers serve as a communication device that allows agents to signal their preferences regarding the final outcome. This in turn can influence the bargaining process by provoking positive or negative emotions and reciprocal behaviour in the other bargainer. In such, offers provide information about one's desires, preferences and bargaining type, which will be judged as more or less kind by the other triggering his reciprocal response. Using a two-stage alternating-offer game over a surplus of \$10, we saw that a responder's counter-proposal is influenced by the initial proposition made by a proposer despite being classified as cheap talk. We found strong support for negative reciprocity as high requests exceeding 60 percent of the surplus are punished partly at own costs by counter-offering an amount which responders expect to be rejected or only just to be accepted by the proposer. The bargaining outcome is critically influenced by the balance of toughness and kindness signalled through the offers made during the haggling phase. Our results indicate that the punishment is not solely driven by distributional preferences but that reciprocal concerns create additional boundaries on how tough one should be in order to reach the best outcome in a bargaining process. The perceived kindness of an ac-

tion in this context depended on the associated belief about the other's intention or type.

Our results help to explain why negotiators avoid adopting an extremely tough bargaining stance in a negotiation, and are well advised to do so as it can come with the cost of a negative reaction from the other bargainer. Yet, this study should be seen as a starting point for further research maybe expanding the number of offer rounds to see how the reciprocal behaviour carries over a number of alterations. It would also be interesting to explore how a larger pie size changes our results. From previous research on the ultimatum game, one might expect a decline in negative reciprocity as the amount at stake increases (e.g. Andersen et al., 2011; Munier and Zaharia, 2002; Slonim and Roth, 1998). Yet, we observe bargaining impasses and delays frequently in everyday-life bargaining situations where high stakes are involved: Especially when people from different cultures with different "bargaining norms" interact, one needs to carefully choose one's bargaining strategy. Not for nothing, there exist many books on how to negotiate with (business) partners from other countries.

The fifth chapter *Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games* experimentally studied the concept of intention-based benevolence by analysing its determinants (in terms of gains and losses created by a first mover's choice) and their interactions. We followed the revealed altruism approach developed by Cox et al. (2008a), and studied how the observable properties of a first mover's choice influence a second mover's decision to be more or less benevolent towards the first mover. With a modified investment game as our work horse, we found that the generosity of a first mover indeed

triggers a reciprocal action as formulated by the reciprocity theory by Cox et al. (2008a). Generosity therefore proved to be a determinant of the other's perceived kindness. However, we found that intention-based benevolence is more than a response to generosity. In particular, we found that the voluntary act of making oneself vulnerable to the other agent's action, i.e. the willingness to take the risk to be worse off after taking a certain action than by maintaining the status quo, also induces pro-social behaviour in a second mover. In this sense, intention-based benevolence is not reducible to reciprocity but (also) a response to the first mover's vulnerability. Generosity is consequently not the only intention triggering pro-social behaviour; vulnerability-responsiveness should receive more attention as a concept on its own. However, while the first mover's generosity and his vulnerability are important drivers of the second mover's benevolence, it is not influenced by the availability of a deal nor an aversion against violating trust (as characterized in the investment game by Berg et al., 1995). Especially the latter finding raises questions regarding the precision of trust-characterization in the trust game; more research emphasis should be put on the features invaluable for trust.

Intention-based benevolence such as trustworthiness and reciprocity play an important role in all our everyday lives. They do not only simplify our social interactions but are often critical for the occurrence of beneficial economic transaction. Especially in situations where contracts are not completely enforceable, standard theory predicts a market breakdown with severe efficiency losses. Thanks to people's justified trust that the other party will not fully exploit their opportunities to disregard their contractual duties, market failures can often be overcome (Fehr and Gächter, 1998). Credence goods markets[1] provide a particu-

[1]The term "credence goods" was first introduced by Darby and Karni (1973) and

larly good example. Despite extreme information asymmetries between sellers (trustees) and consumers (trustors), expert markets such as the health care market or the market for repair services operate – yet not 100 percent efficient. It is therefore crucial to understand the concepts clearly in order to strengthen and upholding trusting relationships from both sides. Our finding of the importance of generosity and vulnerability can thereby be seen as a starting point. However, while the present study focused on a second mover's reaction to a first mover's vulnerability and generosity, the availability of a deal as well as their combined presence, one should note that other factors (that are not present in the trust game) may also be relevant for intention-based benevolence. Besides personality and cultural traits, it would for example be desirable to investigate the role of costs of benevolence. In the traditional investment game and also our design, the second mover's trade-off between own and other's payoff is one-to-one. Increasing efficiency by lowering the costs of benevolence is likely to influence the second mover's pro-social behaviour. But to what extent? And how does efficiency interact with the vulnerability and the generosity of the first mover? Additionally, it would be interesting to investigate whether first movers anticipate second mover's behaviour correctly and adjust their trusting behaviour to the involved own risk/vulnerability as well as the provided benefits towards the other player.

Chapter 6 was devoted to the last study *Guilt-averse or reciprocal: Looking at behavioural motivations in the trust game*. Herein, we experimentally investigated the empirical relevance of the most prominent

---

refers to goods or services where an expert knows more about the quality a customer needs than the consumer himself (Dulleck and Kerschbamer, 2006).

models of belief-dependent motivations for behaviour in the binary trust game. We tested how subjects respond to another player's payoff expectations given a certain distributional outcome. Based on mixed results in previous studies, we hypothesized that the behaviour observed in the different versions of the trust game is generated by heterogeneous mental processes: guilt aversion, kindness-reciprocity, unconditional altruism and own-payoff maximization. In particular, we argued that the effects of the first two are confounded and even masked when looking at averages as their associated behavioural predictions go in opposite directions. Our triadic design implemented within subjects allowed us to study individual response patterns to exogenously manipulated second-order beliefs. Our results however do not support our hypothesis of behavioural types.

Although, in our particular chosen design, heterogeneities seem to be minor, the rejection of the presence of different types may be a overhasty decision. In particular, I want to stress following shortcomings of our experiment: First the number of observations per individual is small, which makes the unambiguous detection of specific patterns difficult. Second, the sample size might not be sufficiently big in order to determine heterogeneities and types accurately and reliably. Moreover and probably most importantly, we are not sure how participants perceive the different continuation probabilities and associated expectations. While we validated our experimental design through the data obtained from impartial observers, we can only say that it works on average. Second-order beliefs may very well vary on the individual level and choices may be responsive to them, yet their inter-dependence will remain unobserved by assuming homogeneous expectations. Closely linked to perceptions is an alternative explanation of our observations. It regards the questions which expectations are decisive for behaviour: conditional or uncondi-

tional expectations? Maybe subjects are more concerned with "choice expectations" (conditional) than with "payoff expectations" (unconditional), which would be an interesting separate research question. A more general concern regards the complexity of the experimental design and the confusion of subjects that may go along with it. The formation of beliefs requires a level of sophistication which is especially high in our set-up as backward induction is required. Without learning opportunities, subjects may simply be cognitively overstrained. These concerns could explain the weakness of our findings on heterogeneous types.[2]

Despite these limitations, our results suggest that the psychological game theories proposed to explain trustworthy behaviour (reciprocity and simple guilt aversion) may not accurately reflect how second movers incorporate their beliefs about the first mover's motive. Instead, other explanations for reciprocal behaviour which are independent of expectations (such as vulnerability-responsiveness or social norms) may play the more important role. Note that the result could for example be explained (at least partly) by the importance of vulnerability that was found in study 2: If first mover's risk to be worse off by trusting is the decisive factor determining second mover's pro-social behaviour, the constant positive back-transfers across decision tasks may be ascribed to an unchanged vulnerability. Further work is needed in this area to understand the role played by perceived intentions in the behaviour of the second mover.

[2]High inter-subject variance can be explained through confusion and would explain the flat slope in the mixture model and insignificant result in the random-coefficient models.

Chapter 8

Discussion

8.1 "Strong reciprocity"?

All the experiments presented in this thesis aimed to investigate people's behaviour in one-shot settings instead of repeated interactions. It might not be far fetched to imagine such one-time encounters nowadays. Yet, in evolutionary terms they were assumingly far less important and the a priori probability of a future reunion was probably never zero. It therefore remains an open question if the observations made really reflect successfully adapted preferences for such conditions (Trivers, 2006). Trivers (2004, 2006) and Binmore (2006) argue that social interactions are intrinsically repeated encounters. The idea is that albeit given full information, individuals do not disassociate themselves completely from the real world where cooperation or pro-social behaviour is typically beneficial. Contrariwise, Bowles and Gintis (2003) as well as Fehr and Schmidt (2006) argue that subjects are aware of the setting they are playing in and are able to leave their real-world experiences of repeated interactions behind. Experimental evidence confirms that individuals cooper-

ate or punish much more in games where they can acquire a reputation or where the probability of meeting again is higher (e.g. Andreoni and Miller, 1993; Gächter and Falk, 2002; Engelmann and Fischbacher, 2009). But is awareness the decisive factor? Trivers (2006) argues that it is irrelevant as humans cannot switch off their biological mechanisms. So, cognitive adjustments might lead to different outcomes in one-shot and repeated games but that does not necessarily mean that the observations made are evolutionary evolved adaptations. In fact, the maximization of fitness does not imply perfect behaviour in every possible situation (West et al., 2011).

The underlying misconception may be a matter of the discrimination between *ultimate* and *proximate* reasons. While proximate explanations are concerned with causal mechanisms underlying behaviour ("how" questions), ultimate explanations regard the resulting fitness consequences of behaviour ("why" questions) (West et al., 2011). Often findings of proximate causes are (misleadingly) accredited to be also the ultimate reason of observations. West et al. (2011) provide an example: Fehr and Fischbacher (2003) describe a proximate mechanism by defining a strong reciprocator as someone who has "a predisposition to reward others for cooperative, norm-abiding behaviours" and "a propensity to impose sanctions on others for norm violations". Nevertheless, they then use it as the solution of the ultimate problem of cooperation: "Strong reciprocity thus constitutes a powerful incentive for cooperation even in non-repeated interactions and when reputation gains are absent". A similar statement can be found in Bowles and Gintis (2004): "[C]ooperation is maintained because many humans have a predisposition to punish those who violate group-beneficial norms". Continuing to ask for the "why" behind reciprocal behaviour, and in particular punishment, quickly leads to an-

other proximate cause: satisfaction (De Quervain et al., 2004). West et al. (2011) argue that this does not solve the ultimate puzzle because it does not answer why evolution should have produced a "psychology or nervous system that mechanistically encourages (rewards) such punishment". The same argument can be made for the "warm glow" after giving (Andreoni, 1990). Although humans may enjoy a warm glow when cooperating, that does not entail that the reason is to generate the warm glow. It may very well be "an unintended by-product" (Roemer, 2015). Indeed, some theorists go as far as suggesting that many of the social emotions evolved after a system of reciprocal altruism had appeared in order to preserve or regulate such social cooperation (Trivers, 1971).

While the debate will continue, the problem of determining whether strong reciprocity provides the ultimate or only the proximate explanation of why humans behave in a certain way in one-shot situations does not derogate our results. In fact, we are actually interested in the proximate motivations of behaviour: How do certain features of and/or beliefs about a decision situation influence an agent's action? We want the data gained in experiments to reflect people's behaviour outside the lab (and not their adjusted behaviour to the stylized and artificial environment). Moreover, it does not mean that we have to give up the picture of the prosocial human who cares about others and their decisions. This is made clear by de Waal and Suchak (2010): Even though aiding behaviour "may very well be evolutionary self-serving (e.g. ultimately increases the actor's fitness through reciprocal altruism or inclusive fitness)", from a proximate perspective it "may be genuinely altruistic in that the actor performs it without selfish ends in mind".

8.2 Trust and trustworthiness

Trusting and trustworthy behaviour is pervading almost all human relationships but also economic transactions (Fehr, 2009). We frequently rely on our friends and family for favours and trust that the baker will hand over the bread rolls after we give him the money. Similarly, we do not drive off after refuelling or leave the restaurant without paying. Few doubt can be raised about the importance of trust and trustworthiness as explanatory factors for social and economic behaviour. In fact, they simplify our life and increase economic welfare for example by saving the costs of writing and policing contracts.

Since the concepts of trust and trustworthiness are so central and relevant in everyday life, they have been the subject of attention in multiple research disciplines including economics, which also entailed its emerging empirical investigation. Some of these studies even suggest a (causal) relationship between trust at the country level and aggregate economic activity such as GDP growth and investment (Knack and Keefer, 1997) or trade volume (Guiso et al., 2009). Kenneth Arrow recognized the importance of trust on economic performance already in 1972 when he wrote: "Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence." (Arrow, 1972, p. 357)

Interestingly, in spite of the importance of these concepts, there exists no consensus about their precise definitions. Bacharach et al. (2007) for instance write: "[D]espite the centrality of trust and trustworthiness in economic activity, and despite the widespread recognition today of their

centrality, there remains much mystification about what produces them, and even about what trust is." Nevertheless, a primary and consistent feature of the existing definitions of trust is vulnerability (Rousseau et al., 1998; Johnson-George and Swap, 1982), prompting (Fehr, 2009, p. 238) to conclude that "the essence of trust [...] consists of the [trustor's] willingness to make herself vulnerable to others' actions".[1] Vulnerability exists if the trustee has an incentive to exploit the trust granted for personal (monetary) gain and thus, it is the risk of being worse off than by not trusting. The definition of trust does not necessarily require a potential gain (or loss) by the trustee although being compatible with it (Ben-Ner and Halldorsson, 2010). The term trustworthiness is mostly used even more vaguely and is less well investigated. Usually, it is used to describe a situation in which the trustee does not fully expropriate the surplus created by the trusting act but shares it with the trustor (e.g. Chaudhuri and Gangadharan, 2007). Alternatively, trustworthiness may mean to not take advantage of the other's vulnerability (Ben-Ner and Halldorsson, 2010). To enhance trust, it is crucial to understand under what circumstances trust is not exploited but honoured by trustworthiness. Analysing the driving factors behind trustworthiness enables people who might trust to better judge the risk involved in their decision.

The investment game (Berg et al., 1995; Fehr and Schmidt, 2006; Johnson and Mislin, 2011) has been the experimental workhorse used by economists to investigate trusting and trustworthy behaviour. The large experimental literature studying this game frequently interpret the observed positive (back-)transfers as trust and trustworthiness, where the latter is on average an increasing function of the exhibited level of trust (Dufwenberg and Gneezy, 2000; Berg et al., 1995; Fehr and Schmidt,

[1] Fehr (2009) base this claim on an article by Hong and Bohnet (2007).

2006). However, Cox (2004) has pointed out that it is unclear whether the investment game is really about trust as it does not eliminate other explanations for the observed behaviour such as unconditional altruism.

In the study *Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games*, we continued the approach taken by Cox (2004) and argue that the observed second mover behaviour in investment games is usually also not separable from *reciprocity*[2]. Specifically, our design allowed us to investigate whether second movers in the investment game behave benevolent as a reaction to the gains and losses created by the trustor's choice – i.e. his generosity or his vulnerability – or because of the possibility of a mutual gain or as a reaction to trust as characterized in the trust game (presence of vulnerability and the possibility of a deal). Our results indicate that the second mover's benevolence is driven by reciprocity and by a vulnerability-responsiveness. We however did not find any evidence that trust itself affects the seond mover's choice beyond the effects of possible gains and losses. Overall, trust as defined in the investment or trust games (i.e. as the combination of vulnerability, generosity and deal-availability) may not be accurate as this specific combination does not enhance benevolence beyond the effects of the former two. A more detailed investigation of the effects of the gains and losses created by a first mover is therefore desirable in order to enhance the understanding of trust as a concept.

Additionally, other factors not present in the trust/investment game (and our experiment) are most likely important drivers of intention-based benevolence (and in particular trustworthiness) – most importantly probably being cultural traits and personality but also the costs of benevo-

[2]Defining "reciprocity" as the combination of rewarding generous behaviour and punishing ungenerous behaviour at own material cost (Cox et al., 2008a; survey: Camerer, 2003b).

lence or the closeness of involved parties.

8.3 Actions, beliefs and perceived intentions

Humans have the ability to theorize about the mental states of others. Having such a "theory of mind" allows us to attribute knowledge, thoughts, beliefs, desires, intentions, and so forth to others which helps to predict or explain others' actions. Furthermore, it enables one to understand that mental states can be the cause of the behaviour of others (Premack and Woodruff, 1978). By constantly hypothesizing about mental states, every human develops a common-sense psychology – ideas about desires and beliefs, and how they influence actions – which is adjusted and fine-tuned in the light of encountered evidence (Suddendorf, 2013, ch. 6). Unheralded flexible scales of social cooperation were probably reached because humans additionally have a fundamental motivation to share their own psychological states with others in order to pursue a shared goal (Tomasello et al., 2005). However, that also means that successful cooperation relies crucially on the exchange of each others' intentions – including attitudes, beliefs, feelings or expectations. In every-day life, there are many social cues that help us to read the other's motivations more accurately. Similarly, we have many means at our disposal in order to let the others know about our own objectives or about how we perceived their actions. Such methods of signalling include facial and body expressions but by far the most powerful instrument is language. Albeit relying on one signal is prone to deception, the combination of several typically gives a fairly good idea about the present intentions.

Unfortunately, in the laboratory setting, most of these channels are hardly ever available to agents. Yet, inter alia the results presented in this thesis show that individuals are often concerned about the other's perceived desires and intentions, and that their decisions depend on them. As intentions typically cannot be observed, players have to rely on actions to form their beliefs about the other's objectives. An open question is, though, *how* players infer intentions and judge behaviour.[3]

I think, one can quickly agree on the necessity of intentionality, in the sense of choice-freedom, for the ascription of an intention to the other person. Blount (1995) and Offerman (2002) provide experimental evidence that deliberate helpful (harmful) choices are rewarded (punished) more frequently than identical but randomly generated choices. One is only (morally and legally) responsible for one's actions if one had a choice and could have acted differently. This argument is also the underlying reason why recent findings in cognitive neuroscience (re-)started a debate in criminal law: A number of studies raise doubts that such a thing as the free will exists. Instead, they suggest that at least some of "our" choices are subconsciously initiated before we become conscious of it (Libet et al., 1983; Haggard, 2011). If criminals are accordingly not responsible for their actions because they are pre-determined by their genetics (and environment), on what legal grounds can we hold them liable and thus prosecute them (e.g. Greene and Cohen, 2004; Jones, 2002; Norrie, 1983)? Although the details of this dispute go far beyond the scope and pertinence of this thesis, the ongoing discussion highlights the relevance and importance of intentionality in other fields as well.

[3]Interestingly, young children's moral judgements and justifications are determined by an action's outcome rather than the actor's intention. Adults, on the other hand, have learned to assign more importance to the aim/objective behind an action than to an action's consequences for moral judgement (Young et al., 2007).

Once intentionality is established, motivations can often be specified through counterfactual reasoning. Interpretations of a certain undertaken action can thereby diverge depending on the available alternatives. What a player could have chosen, but did not, may matter just as much as the choice itself and can thus lead to very different final outcomes of an interaction. Evidence for the relevance of the available actions, the strategy space, is for example shown by Falk et al. (2003): Second movers in the ultimatum game rejected the same offer less often when it is the most generous offer than when it is the least generous in the first movers opportunity set. Different options may thereby influence an action's assessments through their varying characteristics. This is the approach we followed in the essay *Why did he do that? Using counterfactuals to study the effect of intentions in extensive form games* in Chapter 5 where observable features of the available choices were assumed to be the exclusive or at least the dominant determinants of participants' intention-judgements and therefore sufficient to derive their preferences. Since the focus of that study is people's reaction to the specific characteristics investigated, the chosen approach is valid.

That such an evaluation strategy is however not always sufficient can be seen in the essay *Guilt-averse or reciprocal: Looking at behavioural motivations in the trust game* in Chapter 6. We found that, depending on the decision situation, not only the alternatives mattered but that also "external" factors affected subjects' beliefs and perceptions, who altered their choice-interpretation and in turn their responses. In the study, the first movers' options and their decisions were kept constant for second movers throughout all decision tasks. Yet, despite the non-presence of clear patterns, the majority adjusted their behaviour to the exogenously manipulated likelihood to make the choice. As formally modelled by psy-

chological game theory (Geanakoplos et al., 1989), the intention assigned to a certain action depends not only on its observable properties and how they compare to the properties of the option(s) not chosen, but also on a player's beliefs. Perhaps the most influential belief for the intention-judgement is the belief about the other's expectation in regard to one's own choice because it enables inferences about the other's expectations of the final payoff allocation. However, beliefs can be impacted by all sorts of things such as framing (Dufwenberg et al., 2011; Ellingsen et al., 2012) or by communication/actions which can be classified as cheap talk (see Chapter 4). Belief elicitation often seems to be the consequent necessity in order to draw the correct conclusions from observed choices. Yet, the correct belief determination is not trivial and has to be done carefully and in consideration of several drawbacks. A first step to ensure that subjects take the task seriously and report truthfully is making rewards dependent on the accuracy of the stated beliefs. Nevertheless, there are distortions to be considered. In general, the elicitation of beliefs may affect choices but the elicitation of choices may also affect subjects' beliefs (Schotter and Trevino, 2013): Asking subjects about their beliefs may increase their understanding of the game so that they make more sophisticated decisions in line with their beliefs. On the contrary, there exist several arguments on how choices could alter beliefs: Subjects may justify their action (to themselves or the experimenter) either to show that the action was morally acceptable given this belief or that the action was "right", respectively consistent with their belief. The evidence is mixed: some find positive, some negative, and again others find no effects for both interaction effects (for a summary on the results see Schotter and Trevino, 2013). These difficulties probably explain the still reluctant inclusion of beliefs into experimental set-ups. Our strategy to avoid or at

least to minimize the side-effects associated with the belief elicitation was twofold. In the essay *Guilt-averse or reciprocal: Looking at behavioural motivations in the trust game*, we introduced an impartial observer whose average beliefs could be used as instruments for the actual beliefs of interest (second mover's average beliefs). Such a design was however not rich enough for our research question in the essay *Driving a hard bargain is a balancing act: The importance of reciprocity in bargaining*, where we were interested in individual beliefs. Since our main focus was nevertheless on participants' choices, we elicited the beliefs only after we had collected all answers in the choice task. As we did not mention guesses until after choice strategies were made, the belief elicitation should not affect participants' prior decisions. Moreover, I argue that beliefs are neither affected very much (if at all) by the choices made beforehand because subjects had to make relatively many decisions on partitions (due to the use of the strategy method) which lowers the probability of remembering the exact choice previously made.

Summing up, the experimental investigation of intentions is subject to manifold challenges. The interpretation of an action and the derivation of the other's intention in the lab is rather difficult because the auxiliary means like body and vocal language are missing. An action by itself may therefore just not contain enough information to clearly convey/communicate its actor's goal or intention uniquely. As a consequence, it is often crucial to elicit a player's belief about the other's motives and expectations. The assumption of homogeneous beliefs can occasionally be misleading since agents may interpret the same situation differently due to the limited cues.

8.4 Emotions

The idea that many of our decisions are caused by emotions (at least as the proximate cause) rather than being based solely on deliberate considerations can be traced back to the work of David Hume and Adam Smith. John Maynard Keynes (1937) also subscribed to this view: "Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as the result of animal spirits – a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities."[4] Nevertheless, emotions were ignored in economics for a long time. Economic models of decision-making generally assume that agents decide between alternatives (be it goods or actions) by evaluating their desirability and likelihood of their consequences. Decision-makers then aggregate the so obtained information by calculating the resulting expected "utility" and choose the option that expectedly maximizes this utility. Emotions do not per se challenge this conception. And slowly, behavioural and experimental economists are starting to accept the challenge of integrating emotions in their theories.

Expected emotions for instance can rather easily be incorporated in economic models. Expected or anticipated emotions are feelings that are not experienced at the time of decision-making but are foreseen to be undergone in the future (Rick and Loewenstein, 2010). Hence, they are only cognitions of future emotions which arise from thinking about the consequences. Expected emotions can then be modelled just as another characteristic of a certain choice which influences an agent's utility (Rick

[4]The passage can be found in *The General Theory of Employment*, Chapter 12, VII.

and Loewenstein, 2010). Interestingly, Jeremy Bentham (1789) had already assigned a prominent role to emotions in his theory (and discussed their determinants and nature quite extensively) when he first proposed his construct of utility. In fact, he considered utility as the net sum of positive and negative emotions (pleasure and pain). However, the psychological substantiations of utility were subsequently ignored in later models although economists did not explicitly deny the general compatibility of expected emotions with the utility framework (Loewenstein, 2000). Recent examples of the successful incorporation of expected emotions in economic theories include the regret model by Loomes and Sugden (1982) and the theory of guilt aversion by Charness and Dufwenberg (2006).

But not all feelings can be foreseen or accurately predicted. That is why we frequently even delay our decisions, saying "I will see how I feel.". The inclusion of such immediate emotions that are only experienced in the moment of choice is more difficult. Underlying these immediate emotions are often visceral factors which are most often very beneficial in guiding our daily functioning but can also push our decisions in a direction different to the one suggested by a (weighted) long-term cost-benefit analysis (Loewenstein, 2000). Such visceral factors capture people's attention and motivate them to make specific choices or act in a particular way. Visceral responses can thereby change desires rapidly because they are themselves affected by external and/or internal stimuli. Hunger for example can reflect the internal bodily state but appetite can also be (artificially) triggered or magnified by external sources such as the smell of food. The overriding power of these "passions" led David Hume (1739-40) to write the often quoted sentence "[r]eason is, and ought only to be the slave of the passions". However, their fundamentality does not pre-

vent us humans to over-interpret the importance of higher-level cognition for decision-making. Instead, people interpret their own behaviour as a result of deliberate considerations even when this is not the case (see Wegner and Wheatley (1999) for a review on findings). Interestingly, this is true even though immediate emotions often drive one's behaviour against self-interest and people are generally aware of that fact. This does not mean that emotions are irrational in the long-run, though.[5] In the case of moralistic emotions, Trivers (2002) posits that they emerged as strategies in the reciprocity game. Liking for example is the emotion to initiate cooperation, anger protects a person against a cheater, and guilt can prevent a cheater from actually cheating (Trivers, 2002, pp. 39-41; Pinker, 1997, pp. 404-405). In this way, moralistic emotions are an important factor in securing long-term cooperation benefits.

Despite the considerable challenges, some experimental work has been done on the understanding and impact of an immediate emotional and a more deliberate response to moral judgement and subsequent decision-making. Xiao and Houser (2005) for example ask whether one punishes because one feels anger or because the other's action is recognized or interpreted as unkind/uncooperative. They provide responders in the ultimatum game the option to send written messages to proposers in addition to their acceptance/rejection decision. They find that rejection rates drop by half if the offered amount was 20 percent or less of the total. And 80 percent of the messages were expressing "negative emotions". For more generous offers, they found no significant difference in rejection rates. Grimm and Mengel (2011) add that this emotional re-

[5]Evolutionary psychologists argue that the emotional system has evolved in order to carry out fast evaluations of important judgements and decisions that occur repeatedly (Cosmides and Tooby, 2000). Yet, most emotional reactions reflect adapted behaviour to ancient living environments and conditions which have changed significantly over the last centuries. Thus, some emotions may seem irrational nowadays.

action is relatively short-lived. They find that a ten minute delay before responders in the ultimatum game accept or reject reduces the rejection rate of low offers of 20-30 percent of the total surplus from 60-80 to 20 percent.

These findings are especially interesting in the context of the use of the strategy method. Under classical assumption of purely rational agents, emotions do not play a role in decision-making and thus the use of the strategy method remains without consequences. Yet, given the research indicated above, more caution is appropriate. By making conditional decisions for every possible information scenario, participants' emotions are likely to be reduced as their decisions are based on hypothetical considerations (Roth, 1995a). Nevertheless, most studies investigating the effect of using the direct-response versus the strategy method report no qualitative difference in results (e.g. Brandts and Charness, 2011; Zizzo, 2010). Contrariwise, the *level* of punishment (Brandts and Charness, 2011) and trustworthiness (Casari and Cason, 2009) has been shown to be reduced. In such, the presented work is most likely to provide the lower bound of behaviour. The effects of trust, kindness and intention-based benevolence may be larger when emotions are not moderated.

In my view, further research on emotions is needed for attaining a better understanding of the differences and interaction of these two channels of decision-making which Kahneman (2011) calls the fast automatic "System 1" and the slow "System 2" which includes complex computations. More work on the types of emotions experienced and their accurate prediction is also needed, along with the investigation of the circumstances in and the degree to which decisions are guided by expected emotions. In particular, the role of visceral factors underlying immediate emotional responses have to date received too little attention given their enormous

impact on our choices.

Chapter 9

Concluding Remarks

While the papers presented investigate nuances of different motivations underlying pro-social behaviour, the overall picture is clear: People's behaviour differs substantially from the standard economic predictions assuming only self-interested individuals. While one's own material well-being most certainly is an important influence of people's choices, it is definitely not the only one. I would like to conclude by saying that this is lucky for human society and end with a quote by Hirshleifer (1987): "The economist must go beyond the assumption of "economic man" precisely because of the economic advantage of not behaving like economic man – an advantage that presumably explains why the world is not populated solely by economic men."

Bibliography

(2003). The norm of restaurant tipping. *Journal of Economic Behavior & Organization*, 52(3):297–321.

Abreu, D. and Gul, F. (2000). Bargaining and reputation. *Econometrica*, 68(1):85–117.

Al-Ubaydli, O. and Lee, M. S. (2012). Do you reward and punish in the way you think others expect you to? *The Journal of Socio-Economics*, 41(3):336–343.

Anbarci, N., Feltovich, N., and Gürdal, M. Y. (2015). Lying about the price? ultimatum bargaining with messages and imperfectly observed offers. *Journal of Economic Behavior & Organization*, 116:346–360.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., and List, J. A. (2011). Stakes matter in ultimatum games. *American Economic Review*, pages 3427–3439.

Andersson, O., Galizzi, M. M., Hoppe, T., Kranz, S., Van Der Wiel, K., and Wengström, E. (2010). Persuasion in experimental ultimatum games. *Economics Letters*, 108(1):16–18.

Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *Economic Journal*, pages 464–477.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Andreoni, J. and Miller, J. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *Economic Journal*, pages 570–585.

Andreoni, J. and Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, pages 343–362.

Ashraf, N., Bohnet, I., and Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, 9(3):193–208.

Attanasi, G., Battigalli, P., and Nagel, R. (2013). Disclosure of Belief-Dependent Preferences in a Trust Game. Working Papers 506, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Bacharach, M. and Gambetta, D. (2001). Trust in signs. *Trust in society*, 2:148–184.

Bacharach, M., Guerra, G., and Zizzo, D. J. (2007). The self-fulfilling property of trust: An experimental study. *Theory and Decision*, 63(4):349–388.

Balafoutas, L., Fornwagner, H., et al. (2016). The limits of guilt. Technical report.

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT press.

Battigalli, P. and Dufwenberg, M. (2007). Guilt in games. *The American Economic Review*, pages 170–176.

Battigalli, P. and Dufwenberg, M. (2009). Dynamic psychological games. *Journal of Economic Theory*, 144(1):1–35.

Baumeister, R. F., Stillwell, A. M., and Heatherton, T. F. (1994). Guilt: an interpersonal approach. *Psychological bulletin*, 115(2):243.

Bellemare, C., Sebald, A., and Strobel, M. (2011). Measuring the willingness to pay to avoid guilt: estimation using equilibrium and stated belief models. *Journal of Applied Econometrics*, 26(3):437–453.

Bellemare, C., Sebald, A., and Suetens, S. (2015). Heterogeneous guilt aversion and incentive effects.

Ben-Ner, A. and Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, 31(1):64–79.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *The American Economic Review*, 96(5):1652–1678.

Bentham, J. (2007). *An introduction to the principles of morals and legislation*. Courier Corporation. [originally published, 1789].

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142.

Besley, T. and Ghatak, M. (2008). Status incentives. *The American Economic Review*, 98(2):206–211.

Binmore, K. (2005). *Natural justice*. Oxford University Press.

Binmore, K. (2006). Why do people cooperate? *Politics, Philosophy & Economics*, 5(1):81–96.

Binmore, K., Shaked, A., and Sutton, J. (1985). Testing noncooperative bargaining theory: A preliminary study. *American Economic Review*, 75(5):1178–1180.

Blanco, M., Engelmann, D., Koch, A. K., and Normann, H.-T. (2010). Belief elicitation in experiments: is there a hedging problem? *Experimental Economics*, 13(4):412–438.

Blount, S. (1995). When social outcomes aren? t fair: The effect of causal attributions on preferences. *Organizational behavior and human decision processes*, 63(2):131–144.

Bolle, F., Tan, J. H., and Zizzo, D. J. (2014). Vendettas. *American Economic Journal: Microeconomics*, 6(2):93–130.

Bolton, G. E. (1991). A comparative model of bargaining: Theory and evidence. *The American Economic Review*, pages 1096–1136.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *The American Economic Review*, pages 166–193.

Bowles, S. and Gintis, H. (2000). Reciprocity, self-interest, and the welfare state. *Nordic Journal of Political Economy*, 26(1):33–53.

Bowles, S. and Gintis, H. (2003). Origins of human cooperation. *Genetic and cultural evolution of cooperation*, 2003:429–43.

Bowles, S. and Gintis, H. (2004). The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theoretical population biology*, 65(1):17–28.

Bowles, S. and Gintis, H. (2009). Beyond enlightened self-interest: social norms, other-regarding preferences, and cooperative behavior. In *Games, Groups, and the Global Good*, pages 57–78. Springer.

Brandts, J. and Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398.

Burke, M. A. and Young, H. P. (2010). Social norms. *Handbook of Social Economics*, pages 311–36.

Burnham, T. C. (2007). High-testosterone men reject low ultimatum game offers. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1623):2327–2330.

Camerer, C. (2003a). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. (2003b). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F. and Fehr, E. (2004). *Foundations of Human Sociality - Experimental and Ethnograhic Evidence from 15 Small-Scale Societies*, chapter Measuring social norms and preferences using experimental games: A guide for social scientists. Oxford University Press.

Campbell III, C. M. and Kamlani, K. S. (1997). The reasons for wage rigidity: evidence from a survey of firms. *The Quarterly Journal of Economics*, pages 759–789.

Caporael, L. R., Dawes, R. M., Orbell, J. M., and Van de Kragt, A. J. (1989). Selfishness examined: Cooperation in the absence of egoistic incentives. *Behavioral and Brain Sciences*, 12(04):683–699.

Casari, M. and Cason, T. N. (2009). The strategy method lowers measured trustworthy behavior. *Economics Letters*, 103(3):157–159.

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Charness, G. and Kuhn, P. (2011). Lab labor: What can labor economists learn from the lab? *Handbook of labor economics*, 4:229–330.

Charness, G. and Levine, D. I. (2007). Intention and stochastic outcomes: An experimental study. *The Economic Journal*, 117(522):1051–1072.

Charness, G., Masclet, D., and Villeval, M. C. (2010). Competitive preferences and status as an incentive: Experimental evidence. *Groupe d'Analyse et de Théorie Economique working paper*, (1016).

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 117(3):817–869.

Chaudhuri, A. and Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, pages 959–985.

Cialdini, R. B. (2001). *Influence: Science and Practice*. Boston, MA: Allyn and Bacon.

Cialdini, R. B., Trost, M. R., and Newsom, J. T. (1995). Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *Journal of Personality and Social Psychology*, 69(2):318.

Clarke, D. M. (1923). *The Hávamál, with Selections from Other Poems of the Edda, Illustrating the Wisdom of the North in Heathen Times*. Cambridge University Press.

Collard, D. (1975). Edgeworth's propositions on altruism. *The Economic Journal*, 85(338):355–360.

Cooper, D. and Kagel, J. H. (2009). Other regarding preferences: a selective survey of experimental results. *Handbook of experimental economics*, 2.

Cosmides, L. and Tooby, J. (2000). Evolutionary psychology and the emotions. *Handbook of emotions*, 2:91–115.

Cox, J. C. (2004). How to identify trust and reciprocity. *Games and Economic Behavior*, 46(2):260–281.

Cox, J. C. and Deck, C. A. (2005). On the nature of reciprocal motives. *Economic Inquiry*, 43(3):623–635.

Cox, J. C., Friedman, D., and Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59(1):17–45.

Cox, J. C., Friedman, D., and Sadiraj, V. (2008a). Revealed altruism. *Econometrica*, 76(1):31–69.

Cox, J. C. and Hall, D. T. (2010). Trust with private and common property: Effects of stronger property right entitlements. *Games*, 1(4):527–550.

Cox, J. C., Kerschbamer, R., and Neururer, D. (2014). What is trustworthiness and what drives it? Technical report.

Cox, J. C., Sadiraj, K., and Sadiraj, V. (2008b). Implications of trust, fear, and reciprocity for modeling economic behavior. *Experimental Economics*, 11(1):1–24.

Cox, J. C. and Sadiraj, V. (2007). On modeling voluntary contributions to public goods. *Public Finance Review*, 35(2):311–332.

Cox, J. C. and Sadiraj, V. (2012). Direct tests of individual preferences for efficiency and equity. *Economic Inquiry*, 50(4):920–931.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451.

Croson, R., Boles, T., and Murnighan, J. K. (2003). Cheap talk in bargaining experiments: lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, 51(2):143–159.

Darby, M. R. and Karni, E. (1973). Free competition and the optimal amount of fraud. *Journal of law and economics*, pages 67–88.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. Murray, London.

Darwin, C. (1969). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. Murray, London (5th ed.).

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

De Quervain, D. J., Fischbacher, U., Treyer, V., Schellhammer, M., et al. (2004). The neural basis of altruistic punishment. *Science*, 305(5688):1254.

de Waal, F. B. and Suchak, M. (2010). Prosocial primates: selfish and unselfish motivations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2711–2722.

Dufwenberg, M., Gächter, S., and Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, 73(2):459–478.

Dufwenberg, M. and Gneezy, U. (2000). Measuring beliefs in an experimental lost wallet game. *Games and Economic Behavior*, 30(2):163–182.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, 47(2):268–298.

Dulleck, U. and Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic literature*, pages 5–42.

Durham, Y., McKinnon, T., and Schulman, C. (2007). Classroom experiments: Not just fun and games. *Economic Inquiry*, 45(1):162–178.

Ederer, F. and Stremitzer, A. (2014). Promises and expectations.

Edmonds, C. (1855). *Cicero's Three Books of Offices, Or, Moral Duties: Also His Cato Major, an Essay on Old Age ; Laelius, an Essay on Friendship ; Paradoxes ; Scipio's Dream ; And, Letter to Quintus on the Duties of a Magistrate*. Harper's new classical library. Harper & Brothers.

El Mouden, C., Burton-Chellew, M., Gardner, A., and West, S. A. (2012). What do humans maximize? *Evolution and Rationality: Decisions, Co-operation and Strategic Behaviour*, page 23.

Ellingsen, T. and Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory. *American Economic Review*, 98(3):990–1008.

Ellingsen, T., Johannesson, M., Mollerstrom, J., and Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1):117–130.

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107.

Elster, J. (1989). Social norms and economic theory. *The Journal of Economic Perspectives*, 3(4):99–117.

Embrey, M., Fréchette, G. R., and Lehrer, S. F. (2014). Bargaining and reputation: An experiment on bargaining in the presence of behavioural types. *The Review of Economic Studies*, page rdu029.

Engelmann, D. and Fischbacher, U. (2009). Indirect reciprocity and strategic reputation building in an experimental helping game. *Games and Economic Behavior*, 67(2):399–407.

Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4):857–869.

Eriksson, T. and Villeval, M. C. (2012). Respect and relational contracts. *Journal of Economic Behavior & Organization*, 81(1):286–298.

Falk, A., Fehr, E., and Fischbacher, U. (2003). On the nature of fair behavior. *Economic Inquiry*, 41(1):20–26.

Falk, A., Fehr, E., and Fischbacher, U. (2005). Driving forces behind informal sanctions. *Econometrica*, 73(6):2017–2030.

Falk, A., Fehr, E., and Fischbacher, U. (2008). Testing theories of fairness—intentions matter. *Games and Economic Behavior*, 62(1):287–303.

Falk, A. and Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Falk, A. and Heckman, J. J. (2009). Lab experiments are a major source of knowledge in the social sciences. *Science*, 326(5952):535–538.

Fehr, E. (2009). On the economics and biology of trust. *Journal of the European Economic Association*, 7(2-3):235–266.

Fehr, E. and Falk, A. (1999). Wage rigidity in a competitive incomplete contract market. *Journal of Political Economy*, 107(1):106–134.

Fehr, E. and Fischbacher, U. (2002). Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*, 112(478):C1–C33.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, 425(6960):785–791.

Fehr, E. and Gächter, S. (1998). How effective are trust-and reciprocity-based incentives. *Economics, Values, and Organization*, pages 337–363.

Fehr, E. and Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *The Journal of Economic Perspectives*, 14(3):159–181.

Fehr, E., Gächter, S., and Kirchsteiger, G. (1997). Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica*, pages 833–860.

Fehr, E. and Henrich, J. (2003). Is strong reciprocity a maladaptation? on the evolutionary foundations of human altruism.

Fehr, E., Kirchsteiger, G., and Riedl, A. (1993). Does fairness prevent market clearing? an experimental investigation. *The Quarterly Journal of Economics*, 108(2):437–459.

Fehr, E., Naef, M., and Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *The American economic review*, 96(5):1912–1917.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1:615–691.

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual preferences for giving. *The American Economic Review*, pages 1858–1876.

Fleming, P. and Zizzo, D. J. (2015). A simple stress test of experimenter demand effects. *Theory and Decision*, 78(2):219–231.

Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior*, 6(3):347–369.

Gächter, S. and Falk, A. (2002). Reputation and reciprocity: Consequences for the labour relation. *The Scandinavian Journal of Economics*, 104(1):1–26.

Gächter, S., Nosenzo, D., and Sefton, M. (2013). Peer effects in pro-social behavior: Social norms or social preferences? *Journal of the European Economic Association*, 11(3):548–573.

Galinsky, A. D. and Mussweiler, T. (2001). First offers as anchors: the role of perspective-taking and negotiator focus. *Journal of personality and social psychology*, 81(4):657.

Geanakoplos, J., Pearce, D., and Stacchetti, E. (1989). Psychological games and sequential rationality. *Games and Economic Behavior*, 1(1):60–79.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, 206(2):169–179.

Goldreich, D. and Pomorski, Ł. (2011). Initiating bargaining. *The Review of Economic Studies*, 78(4):1299–1328.

Greene, J. and Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philos Trans R Soc Lond B Biol Sci*, 359(1451):1775–85.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Grimm, V. and Mengel, F. (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters*, 111(2):113–115.

Guerra, G. and Zizzo, D. (2004). Trust responsiveness and beliefs. *Journal of Economic Behavior & Organization*, 55(1):25–30.

Guiso, L., Sapienza, P., and Zingales, L. (2009). Cultural biases in economic exchange?. *Quarterly Journal of Economics*, 124(3):1095 – 1131.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4):367–388.

Haggard, P. (2011). Decision time for free will. *Neuron*, 69(3):404–406.

Hamilton, W. D. (1964). The genetical evolution of social behaviour. ii. *Journal of Theoretical Biology*, 7(1):17–52.

Harrison, G. W. and Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, 12(2):133–158.

Hart, O. and Moore, J. (2008). Contracts as reference points. *The Quarterly Journal of Economics*, 123(1):1–48.

Heffetz, O. and Frank, R. H. (2008). Preferences for status: Evidence and economic implications. *Handbook of Social Economics, Jess Benhabib, Alberto Bisin, Matthew Jackson, eds*, 1:69–91.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., et al. (2005). Economic man in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, 28(06):795–815.

Hirshleifer, J. (1987). *On the emotions as guarantors of threats and promises*. Cambridge: MIT Press.

Holt, C. A. (2006). *Markets, games, and strategic behavior: recipes for interactive learning*. Pearson Addison Wesley.

Hong, K. and Bohnet, I. (2007). Status and distrust: The relevance of inequality and betrayal aversion. *Journal of Economic Psychology*, 28(2):197–213.

Hume, D. (2012). *A treatise of human nature*. Courier Corporation. [originally published 1739-40].

Hurley, T. M. and Shogren, J. F. (2005). An experimental comparison of induced and elicited beliefs. *Journal of Risk and Uncertainty*, 30(2):169–188.

Johnson, N. D. and Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology*, 32(5):865–889.

Johnson-George, C. and Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology*, 43(6):1306.

Jones, M. (2002). Overcoming the myth of free will in criminal law: the true impact of the genetic revolution. *Duke LJ*, 52:1031.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of business*, pages S285–S300.

Kawagoe, T. and Narita, Y. (2014). Guilt aversion revisited: An experimental test of a new model. *Journal of Economic Behavior & Organization*, 102:1–9.

Keynes, J. M. (2013). *The general theory of employment, interest and money*. Edison Martin. [originally published 1936].

Knack, S. and Keefer, P. (1997). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly journal of economics*, pages 1251–1288.

Kriss, P. H., Nagel, R., and Weber, R. A. (2013). Implicit vs. explicit deception in ultimatum games with incomplete information. *Journal of Economic Behavior & Organization*, 93:337–346.

Krupka, E., Leider, S., and Jiang, M. (2011). A meeting of the minds: Contracts and social norms. Technical report, mimeo, University of Michigan.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Lazarus, R. S. (1991). *Emotion and adaptation.* Oxford University Press.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, 1(3):593–622.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, pages 153–174.

Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain*, 106(3):623–642.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, pages 426–432.

Loomes, G. and Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, pages 805–824.

McCabe, K. A., Rigdon, M. L., and Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, 52(2):267–275.

Mullainathan, S. and Thaler, R. H. (2000). Behavioral economics. Technical report, National Bureau of Economic Research.

Munier, B. and Zaharia, C. (2002). High stakes and acceptance behavior in ultimatum bargaining. *Theory and Decision*, 53(3):187–207.

Nash, J. (1953). Two-person cooperative games. *Econometrica: Journal of the Econometric Society*, pages 128–140.

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92(1):91–112.

Norrie, A. (1983). Freewill, determinism and criminal justice. *Legal Studies*, 3(1):60–73.

Offerman, T. (2002). Hurting hurts more than helping helps. *European Economic Review*, 46(8):1423–1437.

Okasha, S. (2013). Biological altruism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2013 edition.

Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, 14(3):137–158.

Pennisi, E. (2005). How did cooperative behavior evolve? *Science*, 309(5731):93–93.

Pinker, S. (1997). How the mind works. 1997. *NY: Norton*.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(04):515–526.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 83(5):1281–1302.

Rabin, M. (2002). A perspective on psychology and economics. *European economic review*, 46(4):657–685.

Raiffa, H. (1982). *The art and science of negotiation*. Harvard University Press.

Raiffa, H., Richardson, J., and Metcalfe, D. (2002). *Negotiation analysis: the science and art of collaborative decision making*. Harvard University Press.

Rankin, F. W. (2003). Communication in ultimatum games. *Economics Letters*, 81(2):267–271.

Reuben, E., Sapienza, P., and Zingales, L. (2009). Is mistrust self-fulfilling? *Economics Letters*, 104(2):89–91.

Rick, S. and Loewenstein, G. (2010). The role of emotion in economic behavior. In Lewis, M., Haviland-Jones, J. M., and Barrett, L. F., editors, *Handbook of emotions*, chapter 9. Guilford Press.

Roemer, J. E. (2015).

Ross, L., Greene, D., and House, P. (1977). *Journal of experimental social psychology*, 13(3):279–301.

Roth, A. (1995a). Bargaining experiments. In Kagel, J. and Roth, A., editors, *Handbook of Experimental Economics*. Princeton University Press, Princeton.

Roth, A. E. (1995b). The handbook of experimental economics. chapter Bargaining experiments. Princeton university press Princeton, NJ.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, 23(3):393–404.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica*, pages 97–109.

Sally, D. (1995). Conversation and cooperation in social dilemmas a meta-analysis of experiments from 1958 to 1992. *Rationality and society*, 7(1):58–92.

Samuelson, P. and Nordhaus, W. (1985). Principles of economics.

Schaffner, M. (2013). Technical report, QUT Business School.

Schotter, A. and Sopher, B. (2006). Trust and trustworthiness in games: An experimental study of intergenerational advice. *Experimental Economics*, 9(2):123–145.

Schotter, A. and Trevino, I. (2013). Belief elicitation in the lab. *Annual Review of Economics*, 6(1).

Schweinsberg, M., Ku, G., Wang, C. S., and Pillutla, M. M. (2012). Starting high and ending with nothing: The role of anchors and power in negotiations. *Journal of Experimental Social Psychology*, 48(1):226–231.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, pages 99–118.

Slonim, R. and Roth, A. E. (1998). Learning in high stakes ultimatum games: An experiment in the slovak republic. *Econometrica*, pages 569–596.

Ståhl, I. (1972). *Bargaining Theory*. (Ekonomiska forskningsinstitutet vid Handelshögskolan i Stockholm (EFI)).

Stanca, L. (2009). Measuring indirect reciprocity: Whose back do we scratch? *Journal of Economic Psychology*, 30(2):190–202.

Stanca, L., Bruni, L., and Corazzini, L. (2009). Testing theories of reciprocity: Do motivations matter? *Journal of Economic Behavior & Organization*, 71(2):233–245.

Strassmair, C. (2009). Can intentions spoil the kindness of a gift? an experimental study. Technical report, Munich discussion paper.

Straub, P. G. and Murnighan, J. K. (1995). An experimental investigation of ultimatum games: Information, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior & Organization*, 27(3):345–364.

Suddendorf, T. (2013). *The gap: the science of what separates us from other animals*. Basic Books.

Sugden, R. (1984). Reciprocity: the supply of public goods through voluntary contributions. *Economic Journal*, pages 772–787.

Tingley, D. H. and Walter, B. F. (2011). Can cheap talk deter? an experimental analysis. *Journal of Conflict Resolution*, 55(6):996–1020.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(05):675–691.

Train, K. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Trivers, R. (2002). *Natural selection and social theory: Selected papers of Robert Trivers*. Oxford University Press Oxford.

Trivers, R. (2004). Mutual benefits at all levels of life. *Science*, 304(5673):964–965.

Trivers, R. (2006). Reciprocal altruism: 30 years later. In *Cooperation in primates and humans*, pages 67–83. Springer.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, pages 35–57.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations1. *Econometrica*, 76(6):1467–1480.

Wegner, D. M. and Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54(7):480.

West, S. A., El Mouden, C., and Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior*, 32(4):231–262.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. *Research in experimental economics*, 12:197–292.

Wilcox, N. T. and Feltovich, N. (2000). Thinking like a game theorist: Comment. *University of Houston Department of Economics working paper*.

Wolitzky, A. (2012). Reputational bargaining with minimal knowledge of rationality. *Econometrica*, 80(5):2047–2087.

Xiao, E. and Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20):7398–7401.

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H., and Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of

strong reciprocity. *Proceedings of the National Academy of Sciences*, 109(50):20364–20368.

Young, L., Cushman, F., Hauser, M., and Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, 104(20):8235–8240.

Young, P. (1998). Social norms and economic welfare. *European Economic Review*, 42(3):821–830.

Zizzo, D. J. (2003). Empirical evidence on interdependent preferences: nature or nurture? *Cambridge Journal of Economics*, 27(6):867–880.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1):75–98.

Appendix A

Appendix to Chapter 4

A.1 Instructions – Experiment

General Instructions

General Remarks

Thank you for participating in this experiment on decision-making. During the experiment you and the other participants are asked to make a series of decisions. The money you will earn will depend partly on your own choices and partly on the choices of other participants. All payments will be made confidentially and in cash at the end of the experiment. Please consider all expressions as gender neutral.

Please do not communicate with other participants. If you have any questions after we finish reading the instructions please raise your hand and an experimenter will approach you and answer your question in private.

2 Roles

There are two roles in this experiment: **Player A** and **Player B**. At the start of the experiment you will be assigned to one of these two roles through a randomized procedure. Your role will then remain the same throughout the experiment. Your role will only be known to you. Each Player A will be randomly paired with one Player B. No one will ever be informed about the identity of the participant you were paired with nor will anybody else be informed about the choices you made.

Earnings

You will receive $3 for participating in this experiment. Depending on your decisions and the decisions of other participants you will receive an additional amount according to the rules explained below.

Privacy

This experiment is designed such that nobody, including the experimenters and the other participants, will ever be informed about the choices you or anyone else will make in the experiment. Neither your name nor your student ID will appear on any decision form. The only identifying label will be a number that is known only to you. At the end of the experiment, you are asked one-by-one to collect your earnings in an envelope from a person who has no involvement in and no information about the experiment.

The Game

In this experiment, you play with one other participant. Each Player A will be randomly paired with one Player B. Each A/B pair can divide an amount of $10 among themselves.

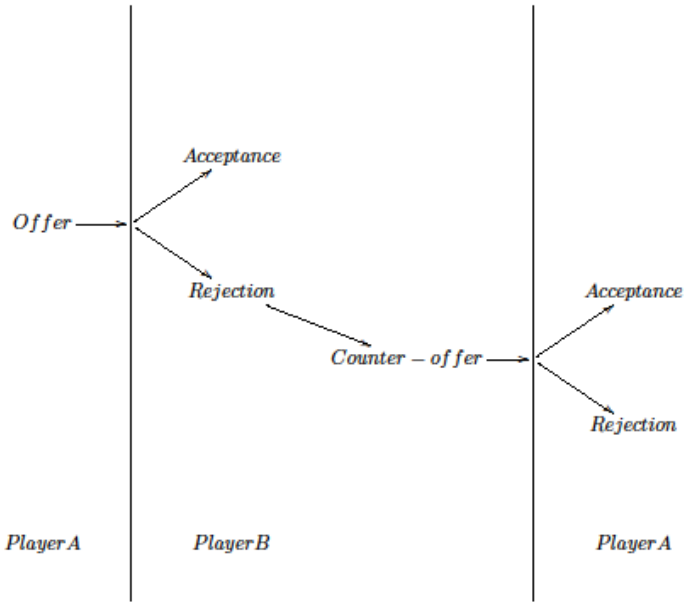
1. Player A proposes how he thinks the $10 should be divided between him and Player B.

2. Player B can then either "accept" or "reject" Player A's proposition.

    - If he accepts, both players will receive an amount according to Player A's suggested partition.
    - If he rejects Player A's proposition, Player B makes a counteroffer regarding the split of the $10 between him and Player A.

3. If Player B has rejected and made his counteroffer, Player A can either "accept" or "reject" the partition of the $10 that is proposed to him by Player B.

    - If Player A accepts, both players will receive an amount according to Player B's suggested partition.
    - If Player A rejects, both players earn nothing.



Decision Tasks

Decision Task - Player A

If you are assigned the role of Player A, you will make two decisions:

1. You are asked to make a proposition on how much of the $10 you want to keep for yourself**.** You can keep any amount between $0 and $10. The rest (if any) will be offered to Player B.
2. You will not be informed whether Player B accepts or rejects your proposition before the end of the experiment. You are therefore asked to state the minimum amount you would need to receive in a counteroffer to just accept it. This means that you would accept all counteroffers above or equal to this amount and reject all below it.

Decision Task - Player B

If you are assigned the role of Player B, you are asked to decide whether you accept or reject Player A's offer regarding the split of the $10.

If you reject his offer, you are asked to make a counteroffer as to how you think the $10 should be divided. In your counteroffer, you can keep any amount between $0 and $10. The rest (if any) will be offered to Player A.

During the experiment you will not be informed about Player A's actual offer. You therefore make all your choices for each possible offer from Player A.

Earnings

At the end of the experiment the cash payments are determined for each pair of participants:

		Payoff Player A	**Payoff Player B**
a) **B accepts** A's offer of $X		$10 - $X	$X
b) **B rejects** A's offer of $X and	• B's counteroffer of $Y for A is bigger than (or equal to) A's minimum amount to accept	$Y	$10 - $Y
	• B's counteroffer of $Y for A is smaller than A's minimum amount to accept	$0	$0

Instructions - Part II

We ask you now to guess some of the answers the other player gave or is about to give. You can earn additional money for your guesses if they are correct.

Player A -- Tasks

If you were assigned the role of Player A, we ask you to **guess if Player B will accept or reject your offer** for all offers that you could have made. In case you think he will reject it, we also ask you to **guess how much Player B will offer you** in his counteroffer.

Player B -- Tasks

If you were assigned the role of Player B, we ask you to make two guesses about answers Player A gave or is about to give:

1. We asked Player A to guess your reaction to his proposed split of the $10. We now ask you to **guess how Player A thinks you would react** for each of his potential partition propositions: If Player A offered you $X, does he expect you to accept his offer? And if you think he does not expect you to accept it, how much does he think you will offer him in your counteroffer?

2. Suppose you rejected Player A's proposition. We ask you to **guess how much you have to give at least to Player A in your counteroffer so that he still accepts it**. This means that you would expect Player A to accept all of your counteroffers, in which you give him more than $X, and to reject all of your counteroffers, in which you give him less than $X. We ask you to make a guess about $X for each of Player A's initial offers.

Earnings

You can earn additional money if your stated guesses match the actual answers given by the other player: You earn **$0.50 for every correct guess**.

- **Player A**: You receive an additional $0.50 for each correct belief about Player B's acceptance decision. In case you correctly guess that Player B rejects your offer, you earn extra $0.50 if your guess about Player B's counteroffer coincides with his actual offer.

- **Player B**: You receive additional money if your guess about Player A's expectation of your reaction matches his actual expectation. Here you can earn $0.50 for each correct guess on whether Player A expects you to accept his offer and in case of a correctly expected rejection you receive $0.50 for each correct guess about what he expects you to counteroffer him.
  Furthermore, you earn an additional $0.50 for each correct guess about Player A's smallest accepted counteroffer.

A.2 Instructions – Questionnaire

General Instructions

General Remarks

Thank you for participating in this experiment on decision-making. During the experiment you and the other participants are asked to answer a series of questions. At the end you will receive a flat payment of **$10**, which is independent of the answers you will give. Please, nevertheless, read the instructions carefully and answer the questions truthfully.

Please consider all expressions as gender neutral.

Please do not communicate with other participants. If you have any questions after we finish reading the instructions please raise your hand and an experimenter will approach you and answer your question in private.

Your Role

In this experiment, you are asked to take the role of an **impartial Observer** whose task is to make guesses and judgements concerning the behaviour of players in the game described below. It is a typical game that has been played out many times.

Privacy

This experiment is designed such that nobody, including the experimenters and the other participants, will ever be informed about the answers you or anyone else will give in the experiment. Neither your name nor your student ID will appear on any decision form. The only identifying label will be a number that is known only to you..

Your Questionnaire

Suppose Player A's initial proposition was to keep **$3 for himself** and to offer **$7 to Player B**.

1. Do you think Player A expects Player B to accept or to reject his proposition?

   ☐ A expects B to accept.
   ☐ A expects B to reject.

2. Now, suppose that B rejects A's initial proposition to keep $8 for himself. What do you think is the lowest counter-offer Player A would then accept from Player B? This means that you would expect such a Player A to accept all counter-offers, in which he receives more or equal than …

   ☐ $0
   ☐ $1
   ☐ $2
   ☐ $3
   ☐ $4
   ☐ $5
   ☐ $6
   ☐ $7
   ☐ $8
   ☐ $9
   ☐ $10

   … and to reject all counter-offers, in which he gets less than the amount.

3. How much do you think a Player A who makes such a proposition expects to earn at the end?

   ☐ $0
   ☐ $1
   ☐ $2
   ☐ $3
   ☐ $4
   ☐ $5
   ☐ $6
   ☐ $7
   ☐ $8
   ☐ $9
   ☐ $10

4. On a scale from 0 to 10, how **fair** do you think Player A's proposition is?

0 – very unfair	1	2	3	4	5	6	7	8	9	10 – very fair
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

5. On a scale from 0 to 10, how **kind** do you think Player A's proposition is?

0 – very unkind	1	2	3	4	5	6	7	8	9	10 – very kind
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

6. Why do you think Player A made a proposition, in which he keeps $8? (You can cross more than one answer.)

   - ☐ He is tough / would reject low counter-offers.
   - ☐ He wants to appear tough / make Player A think that he would reject low counter-offers.
   - ☐ He is fair.
   - ☐ He is smart.
   - ☐ He wants to assure himself the larger part of the $10.
   - ☐ He is kind.
   - ☐ He is selfish.
   - ☐ He is nasty.
   - ☐ He did not understand the game.
   - ☐ Other:

     _____

     _____

     _____

     _____

7. Receiving such an offer, in which you get $2 of the $10, how would you feel? (You can cross more than one answer.)

   - ☐ Fine.
   - ☐ Insulted.
   - ☐ Understanding.
   - ☐ Good.
   - ☐ Neutral.
   - ☐ Angry.
   - ☐ Happy.
   - ☐ Other:

     _____

     _____

     _____

     _____

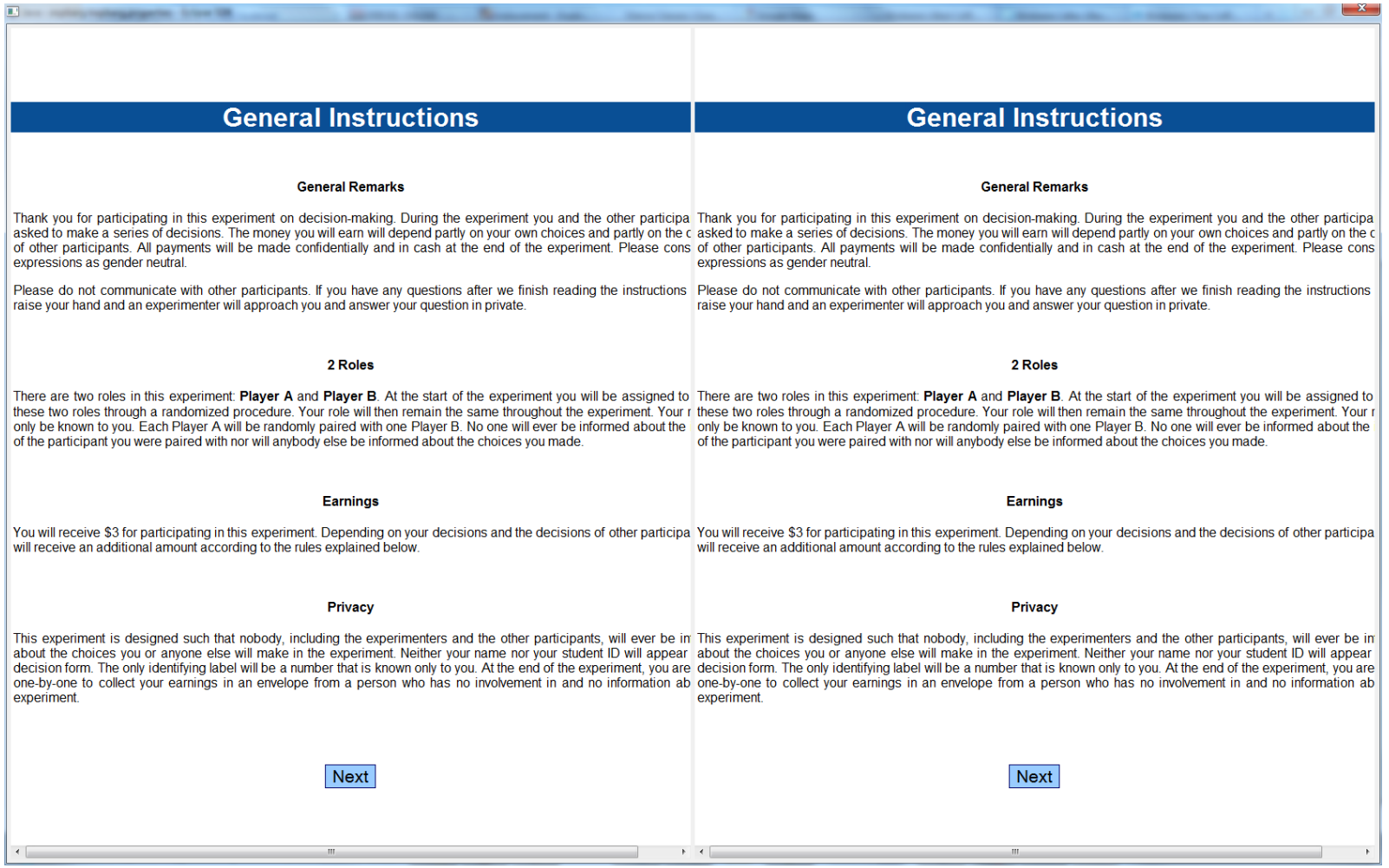
8. After receiving such a proposal, in which you receive $2 of the $10, would you consider to make a counter-offer to A, which is so low that you find it likely to be rejected by A?

   - ☐ Yes.
   - ☐ No.

9. Suppose you were Player B and had the choice to end the game without an agreement (payoff of $0 for both of you). After receiving such a proposal, in which you receive $2 of the $10, would you choose to end the game?

   - ☐ Yes.
   - ☐ No.

A.3 Screenshots – Experiment



General Instructions

General Remarks

Thank you for participating in this experiment on decision-making. During the experiment you and the other participa asked to make a series of decisions. The money you will earn will depend partly on your own choices and partly on the c of other participants. All payments will be made confidentially and in cash at the end of the experiment. Please cons expressions as gender neutral.

Please do not communicate with other participants. If you have any questions after we finish reading the instructions raise your hand and an experimenter will approach you and answer your question in private.

2 Roles

There are two roles in this experiment: **Player A** and **Player B**. At the start of the experiment you will be assigned to these two roles through a randomized procedure. Your role will then remain the same throughout the experiment. Your r only be known to you. Each Player A will be randomly paired with one Player B. No one will ever be informed about the of the participant you were paired with nor will anybody else be informed about the choices you made.

Earnings

You will receive $3 for participating in this experiment. Depending on your decisions and the decisions of other participa will receive an additional amount according to the rules explained below.

Privacy

This experiment is designed such that nobody, including the experimenters and the other participants, will ever be in about the choices you or anyone else will make in the experiment. Neither your name nor your student ID will appear decision form. The only identifying label will be a number that is known only to you. At the end of the experiment, you are one-by-one to collect your earnings in an envelope from a person who has no involvement in and no information ab experiment.

Next

General Instructions

General Remarks

Thank you for participating in this experiment on decision-making. During the experiment you and the other participa asked to make a series of decisions. The money you will earn will depend partly on your own choices and partly on the c of other participants. All payments will be made confidentially and in cash at the end of the experiment. Please cons expressions as gender neutral.

Please do not communicate with other participants. If you have any questions after we finish reading the instructions raise your hand and an experimenter will approach you and answer your question in private.

2 Roles

There are two roles in this experiment: Player A and Player B. At the start of the experiment you will be assigned to these two roles through a randomized procedure. Your role will then remain the same throughout the experiment. Your r only be known to you. Each Player A will be randomly paired with one Player B. No one will ever be informed about the of the participant you were paired with nor will anybody else be informed about the choices you made.

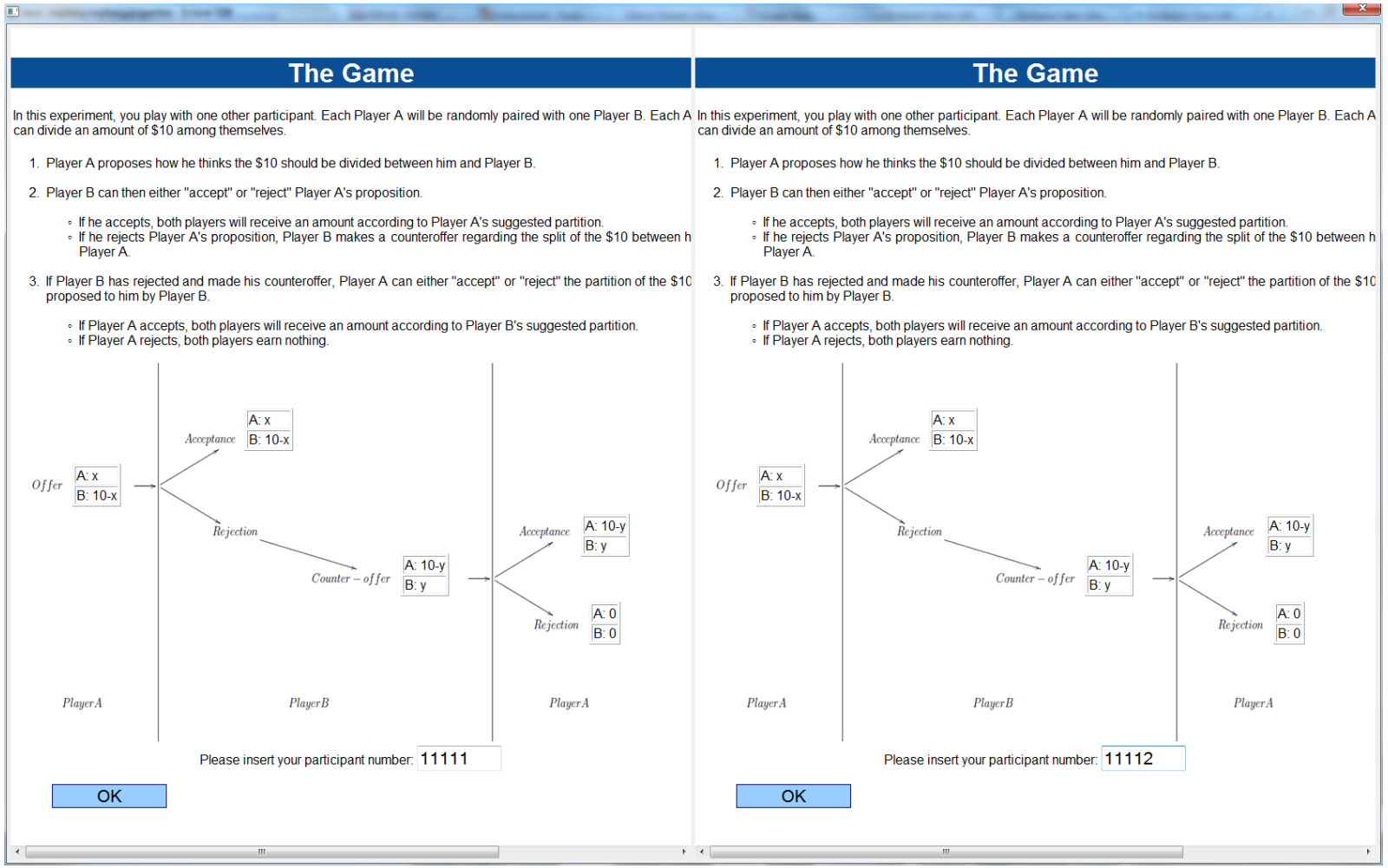
Earnings

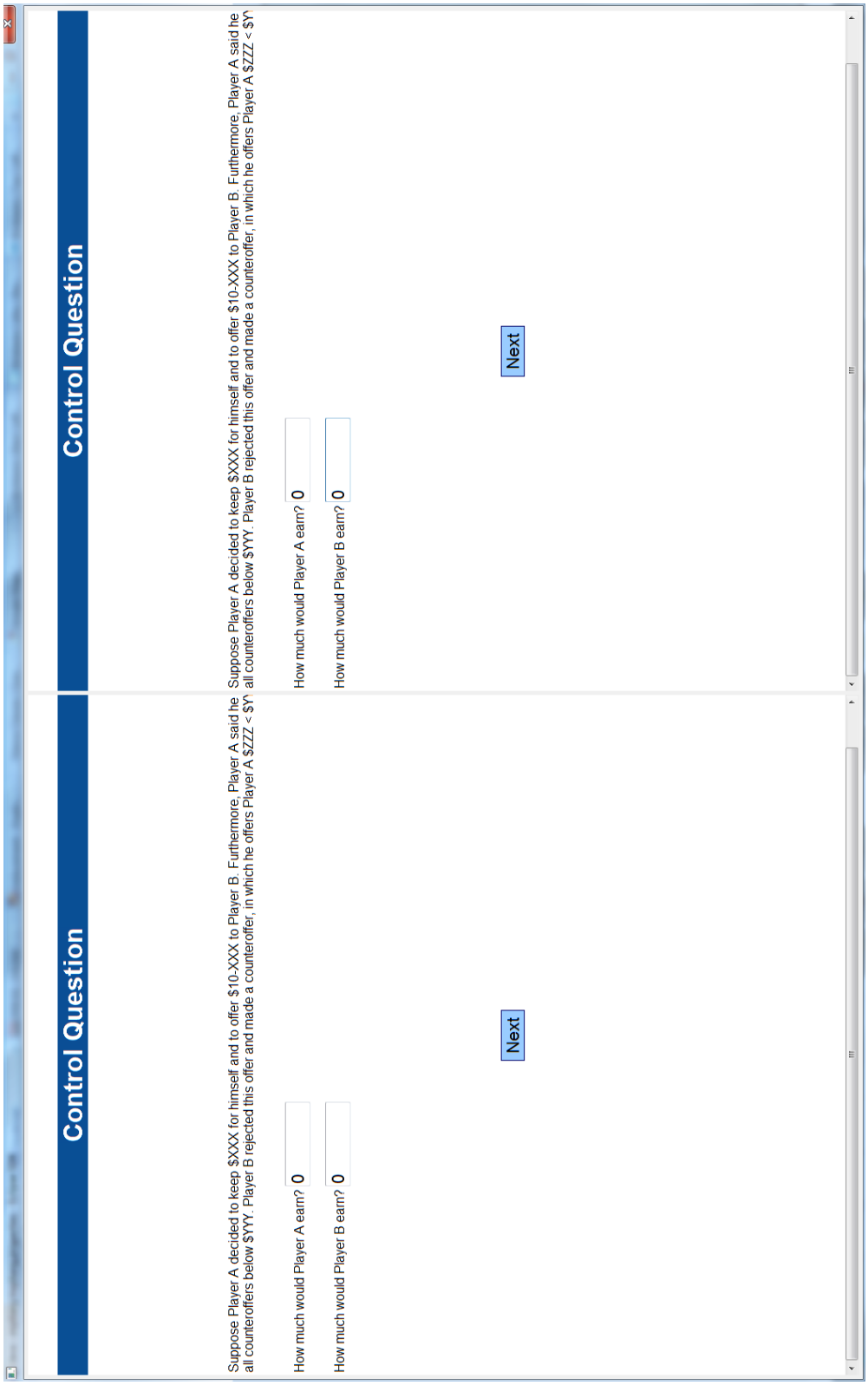
You will receive $3 for participating in this experiment. Depending on your decisions and the decisions of other participa will receive an additional amount according to the rules explained below.

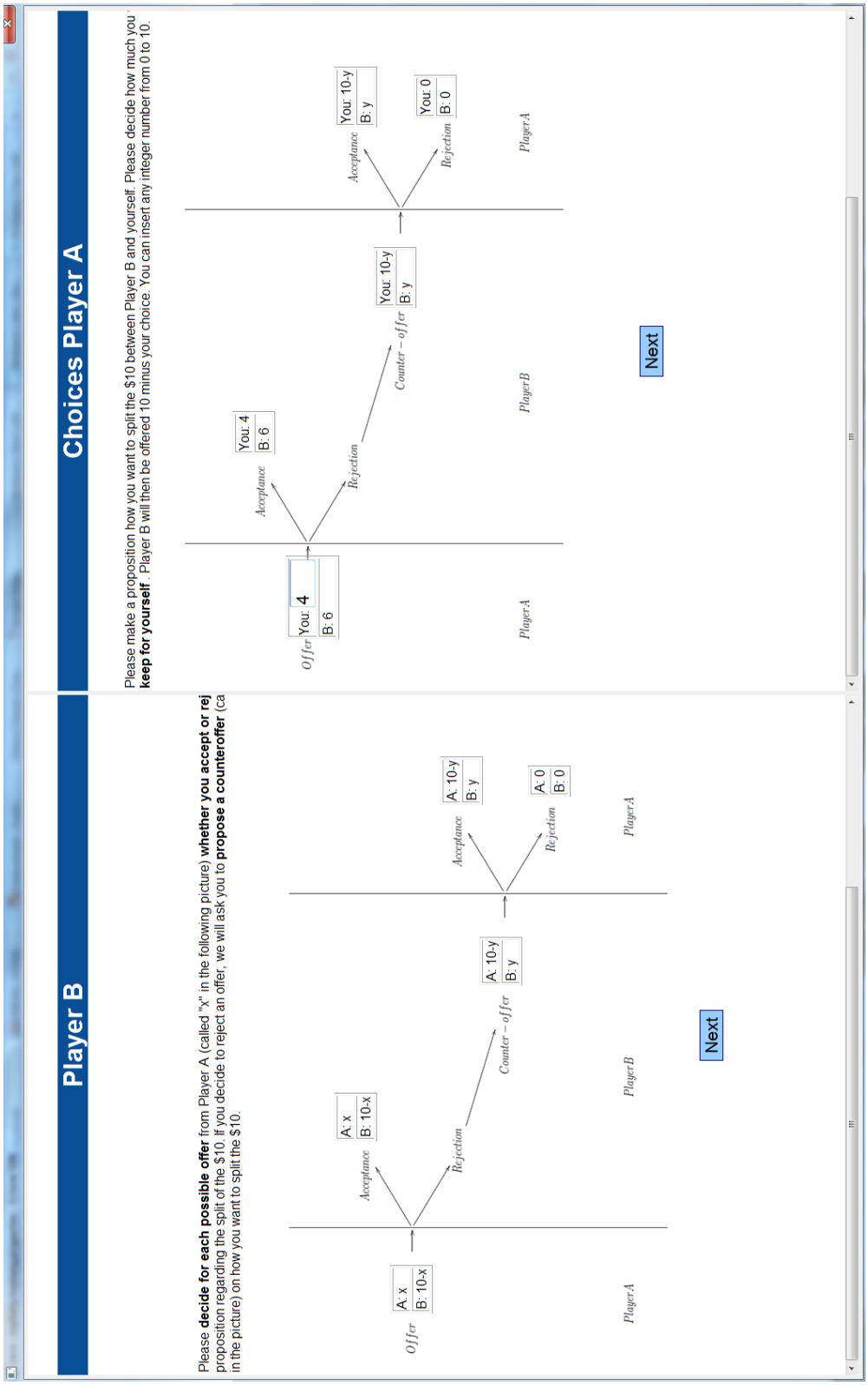
Privacy

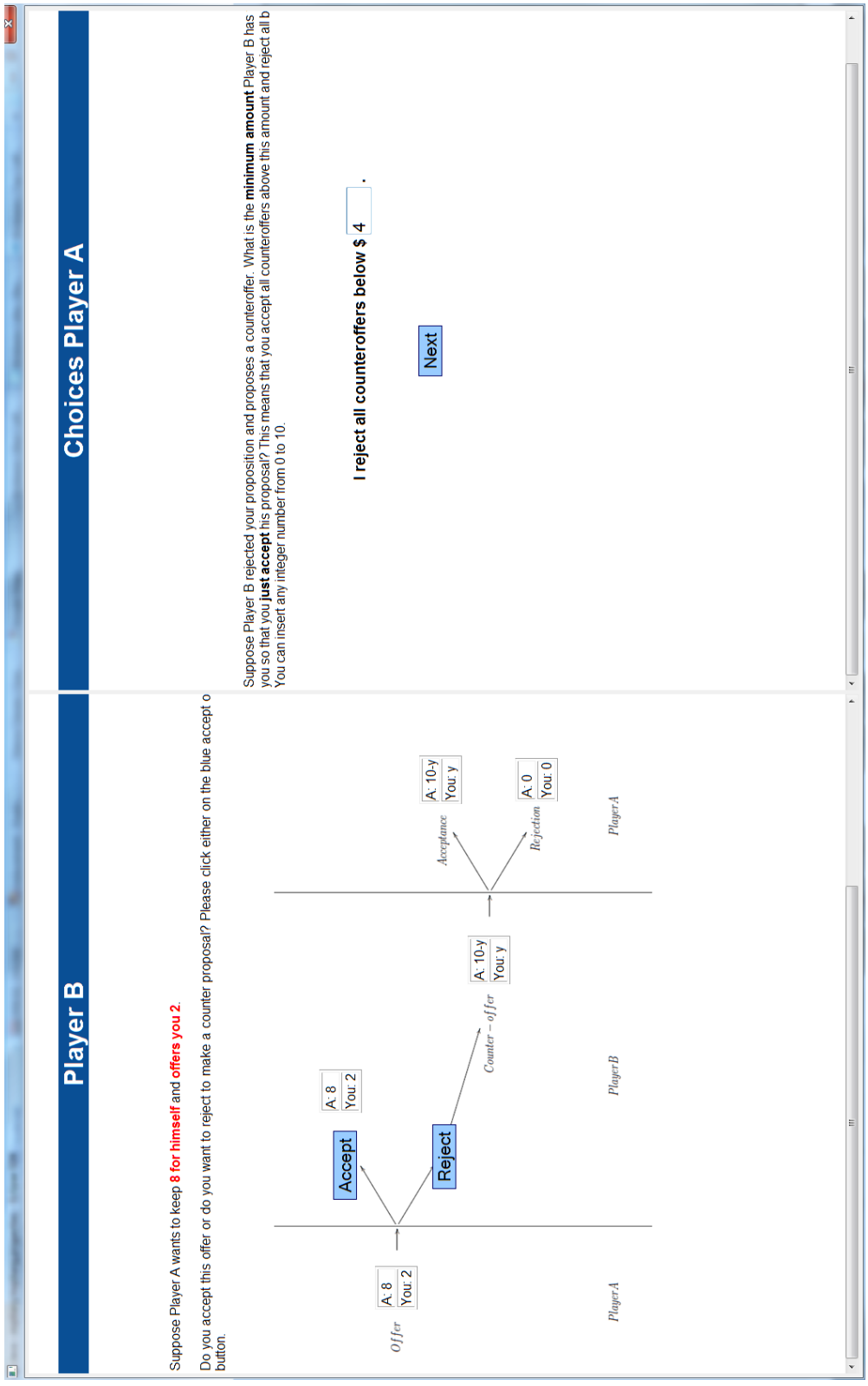
This experiment is designed such that nobody, including the experimenters and the other participants, will ever be in about the choices you or anyone else will make in the experiment. Neither your name nor your student ID will appear decision form. The only identifying label will be a number that is known only to you. At the end of the experiment, you are one-by-one to collect your earnings in an envelope from a person who has no involvement in and no information ab experiment.

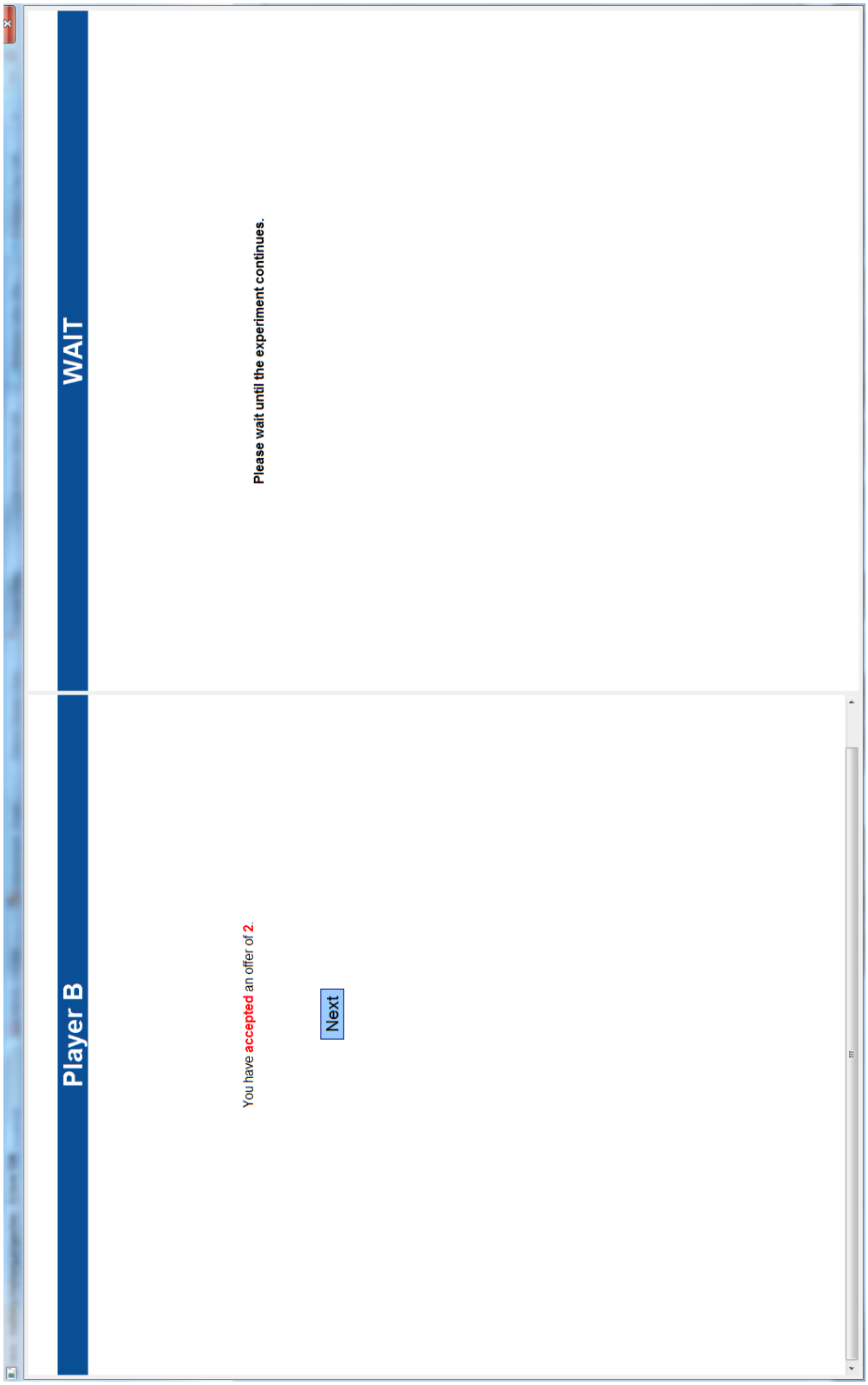
Next

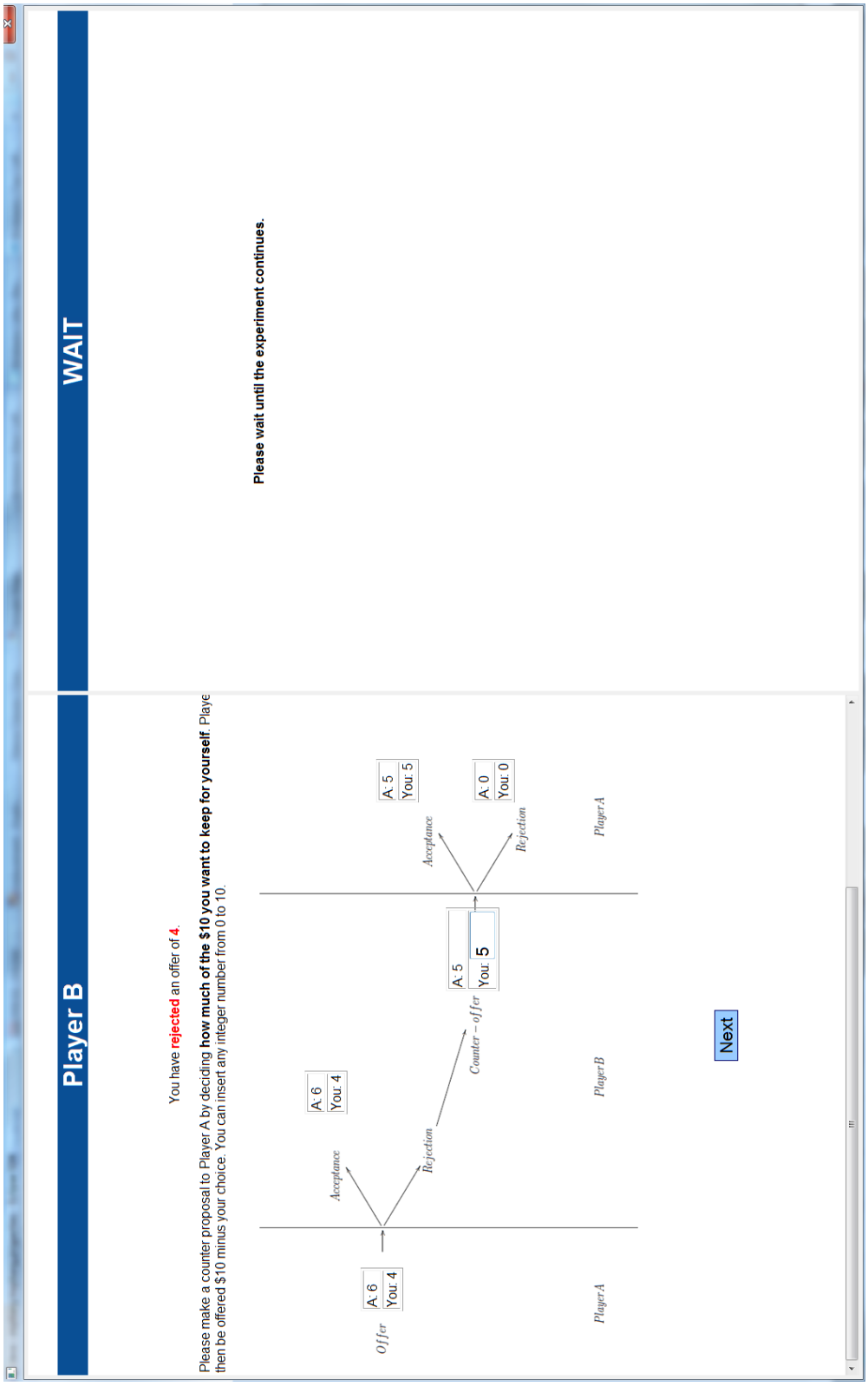

The Game
In this experiment, you play with one other participant. Each Player A will be randomly paired with one Player B. Each A can divide an amount of $10 among themselves.
1. Player A proposes how he thinks the $10 should be divided between him and Player B.
2. Player B can then either "accept" or "reject" Player A's proposition.
◦ If he accepts, both players will receive an amount according to Player A's suggested partition.
◦ If he rejects Player A's proposition, Player B makes a counteroffer regarding the split of the $10 between h Player A.
3. If Player B has rejected and made his counteroffer, Player A can either "accept" or "reject" the partition of the $10 proposed to him by Player B.
◦ If Player A accepts, both players will receive an amount according to Player B's suggested partition.
◦ If Player A rejects, both players earn nothing.
Offer
A: x
B: 10-x
Acceptance
A: x
B: 10-x
Rejection
Counter – offer
A: 10-y
B: y
Acceptance
A: 10-y
B: y
Rejection
A: 0
B: 0
PlayerA
PlayerB
PlayerA
Please insert your participant number: 11111
OK
The Game
In this experiment, you play with one other participant. Each Player A will be randomly paired with one Player B. Each A can divide an amount of $10 among themselves.
1. Player A proposes how he thinks the $10 should be divided between him and Player B.
2. Player B can then either "accept" or "reject" Player A's proposition.
◦ If he accepts, both players will receive an amount according to Player A's suggested partition.
◦ If he rejects Player A's proposition, Player B makes a counteroffer regarding the split of the $10 between h Player A.
3. If Player B has rejected and made his counteroffer, Player A can either "accept" or "reject" the partition of the $10 proposed to him by Player B.
◦ If Player A accepts, both players will receive an amount according to Player B's suggested partition.
◦ If Player A rejects, both players earn nothing.
Offer
A: x
B: 10-x
Acceptance
A: x
B: 10-x
Rejection
Counter – offer
A: 10-y
B: y
Acceptance
A: 10-y
B: y
Rejection
A: 0
B: 0
PlayerA
PlayerB
PlayerA
Please insert your participant number: 11112
OK


Control Question
Suppose Player A decided to keep $XXX for himself and to offer $10-XXX to Player B. Furthermore, Player A said he all counteroffers below $YYY. Player B rejected this offer and made a counteroffer, in which he offers Player A $ZZZ < $Y
How much would Player A earn? 0
How much would Player B earn? 0
Next
Control Question
Suppose Player A decided to keep $XXX for himself and to offer $10-XXX to Player B. Furthermore, Player A said he all counteroffers below $YYY. Player B rejected this offer and made a counteroffer, in which he offers Player A $ZZZ < $Y
How much would Player A earn? 0
How much would Player B earn? 0
Next


Choices Player A
Please make a proposition how you want to split the $10 between Player B and yourself. Please decide how much you
keep for yourself. Player B will then be offered 10 minus your choice. You can insert any integer number from 0 to 10.
Offer
You: 4
B: 6
Player A
Acceptance
You: 4
B: 6
Rejection
Player B
Counter – offer
You: 10-y
B: y
Player A
Acceptance
You: 10-y
B: y
Rejection
You: 0
B: 0
Player A
Next
Player B
Please decide for each possible offer from Player A (called "x" in the following picture) whether you accept or rej
proposition regarding the split of the $10. If you decide to reject an offer, we will ask you to propose a counteroffer (ca
in the picture) on how you want to split the $10.
Offer
A: x
B: 10-x
Player A
Acceptance
A: x
B: 10-x
Rejection
Player B
Counter – offer
A: 10-y
B: y
Player A
Acceptance
A: 10-y
B: y
Rejection
A: 0
B: 0
Player A
Next


Player B
Choices Player A
Suppose Player A wants to keep 8 for himself and offers you 2.
Do you accept this offer or do you want to reject to make a counter proposal? Please click either on the blue accept o
button.
Suppose Player B rejected your proposition and proposes a counteroffer. What is the minimum amount Player B has
you so that you just accept his proposal? This means that you accept all counteroffers above this amount and reject all b
You can insert any integer number from 0 to 10.
I reject all counteroffers below $ 4 .
Next
Offer
A: 8
You: 2
Accept
Reject
A: 8
You: 2
Counter – offer
A: 10-y
You: y
Acceptance
A: 10-y
You: y
Rejection
A: 0
You: 0
Player A
Player B
Player A


WAIT
Please wait until the experiment continues.
Player B
You have accepted an offer of 2.
Next


Player B
WAIT
You have rejected an offer of 4.
Please make a counter proposal to Player A by deciding how much of the $10 you want to keep for yourself. Player
then be offered $10 minus your choice. You can insert any integer number from 0 to 10.
Please wait until the experiment continues.
Offer
A: 6
You: 4
Player A
Acceptance
A: 6
You: 4
Rejection
Player B
Counter – offer
A: 5
You: 5
Player A
Acceptance
A: 5
You: 5
Rejection
A: 0
You: 0
Next

Instructions - Part II

We ask you now to guess some of the answers the other player gave or is about to give. You can earn additional money f guesses if they are correct.

Player A -- Tasks

If you were assigned the role of Player A, we ask you to **guess if Player B will accept or reject your offer** for all offe you could have made. In case you think he will reject it, we also ask you to **guess how much Player B will offer you** counteroffer.

Player B -- Tasks

If you were assigned the role of Player B, we ask you to make two guesses about answers Player A gave or is about to g

1. We asked Player A to guess your reaction to his proposed split of the $10. We now ask you to **guess how Pl thinks you would react** for each of his potential partition propositions. If Player A offered you $X, does he expec accept his offer? And if you think he does not expect you to accept it, how much does he think you will offer him counteroffer?
2. Suppose you rejected Player A's proposition. We ask you to **guess how much you have to give at least to Pl in your counteroffer so that he still accepts it**. This means that you would expect Player A to accept all counteroffers, in which you give him more than $X, and to reject all of your counteroffers, in which you give him le $X. We ask you to make a guess about $X for each of Player A's initial offers.

Earnings

You can earn additional money if your stated guesses match the actual answers given by the other player. You earn **$0 every correct guess**.

- **Player A**: You receive an additional $0.50 for each correct belief about Player B's acceptance decision. In ca correctly guess that Player B rejects your offer, you earn extra $0.50 if your guess about Player B's counteroffer coi with his actual offer.
- **Player B**: You receive additional money if your guess about Player A's expectation of your reaction matches his expectation. Here you can earn $0.50 for each correct guess on whether Player A expects you to accept his offer case of a correctly expected rejection you receive $0.50 for each correct guess about what he expects you to coun him.
Furthermore, you earn an additional $0.50 for each correct guess about Player A's smallest accepted counteroffer.

Next

Instructions - Part II

We ask you now to guess some of the answers the other player gave or is about to give. You can earn additional money f guesses if they are correct.

Player A -- Tasks

If you were assigned the role of Player A, we ask you to guess if Player B will accept or reject your offer for all offe you could have made. In case you think he will reject it, we also ask you to guess how much Player B will offer you counteroffer.

Player B -- Tasks

If you were assigned the role of Player B, we ask you to make two guesses about answers Player A gave or is about to g

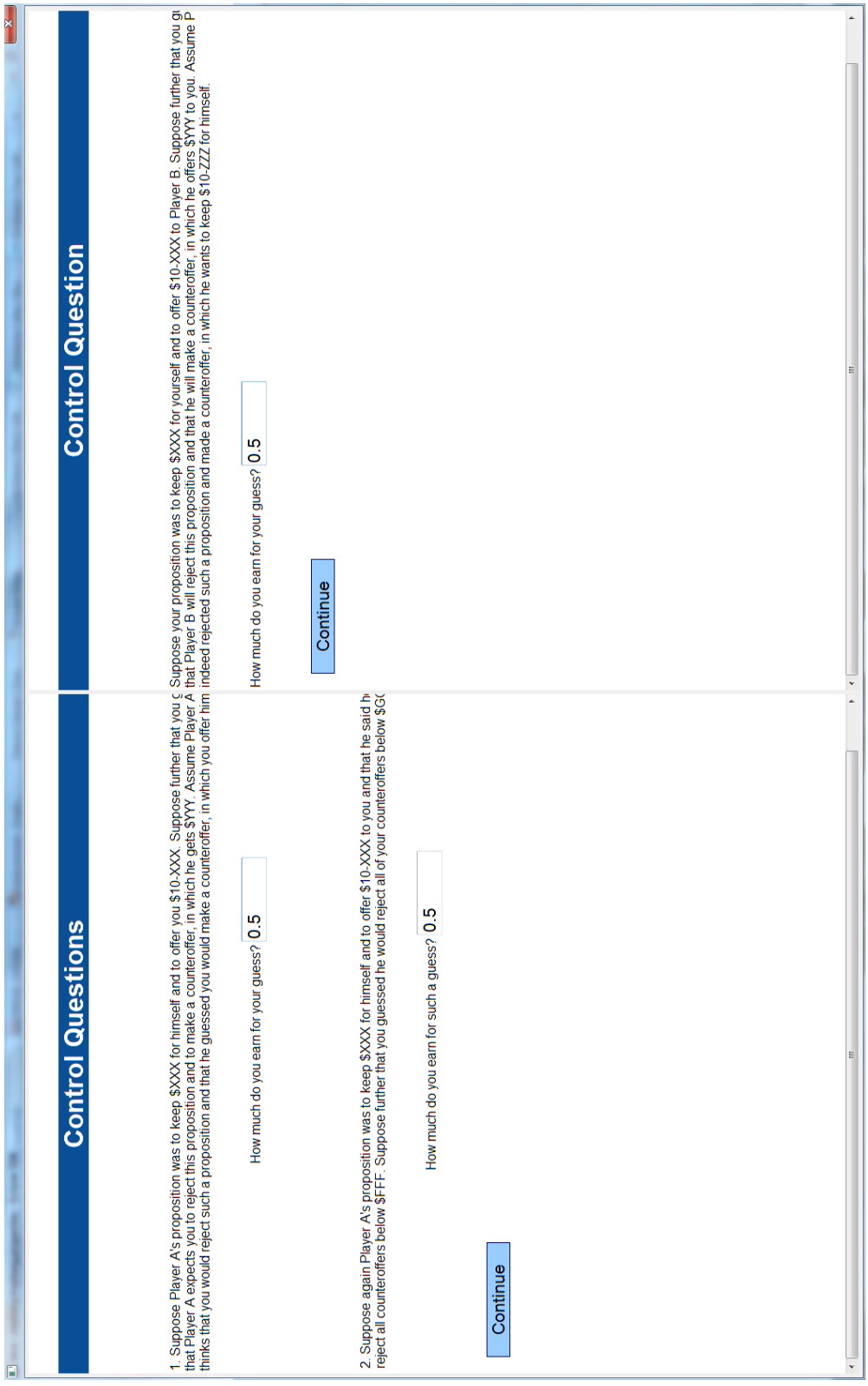
1. We asked Player A to guess your reaction to his proposed split of the $10. We now ask you to guess how Pl thinks you would react for each of his potential partition propositions. If Player A offered you $X, does he expec accept his offer? And if you think he does not expect you to accept it, how much does he think you will offer him counteroffer?
2. Suppose you rejected Player A's proposition. We ask you to guess how much you have to give at least to Pl in your counteroffer so that he still accepts it. This means that you would expect Player A to accept all counteroffers, in which you give him more than $X, and to reject all of your counteroffers, in which you give him le $X. We ask you to make a guess about $X for each of Player A's initial offers.

Earnings

You can earn additional money if your stated guesses match the actual answers given by the other player. You earn $0 every correct guess.

- Player A: You receive an additional $0.50 for each correct belief about Player B's acceptance decision. In ca correctly guess that Player B rejects your offer, you earn extra $0.50 if your guess about Player B's counteroffer coi with his actual offer.
- Player B: You receive additional money if your guess about Player A's expectation of your reaction matches his expectation. Here you can earn $0.50 for each correct guess on whether Player A expects you to accept his offer case of a correctly expected rejection you receive $0.50 for each correct guess about what he expects you to coun him.
Furthermore, you earn an additional $0.50 for each correct guess about Player A's smallest accepted counteroffer.

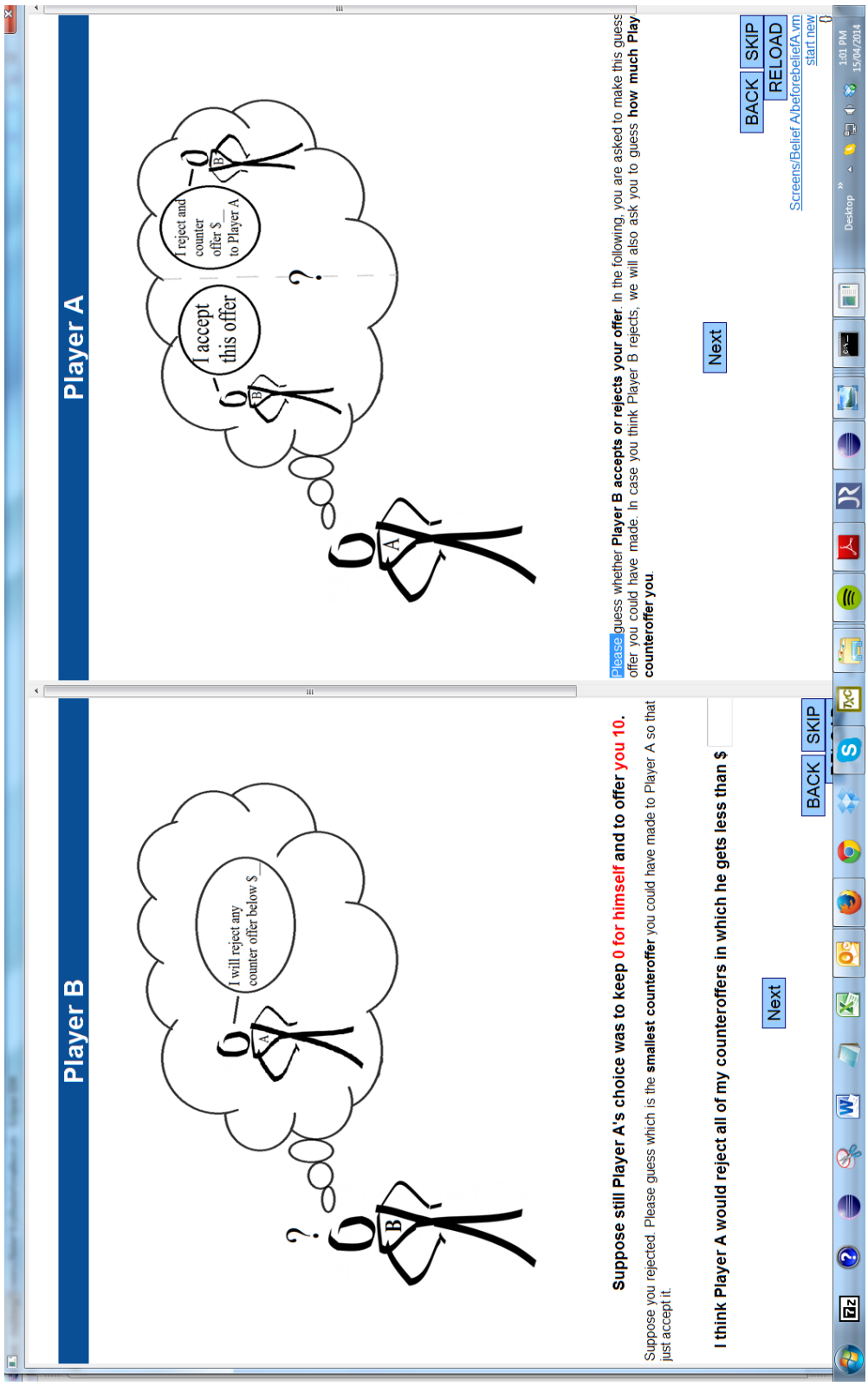
Next


Control Questions
1. Suppose Player A's proposition was to keep $XXX for himself and to offer you $10-XXX. Suppose further that you
that Player A expects you to reject this proposition and to make a counteroffer, in which he gets $YYY. Assume Player A
thinks that you would reject such a proposition and that he guessed you would make a counteroffer, in which you offer him
How much do you earn for your guess? 0.5
2. Suppose again Player A's proposition was to keep $XXX for himself and to offer $10-XXX to you and that he said h
reject all counteroffers below $FFF. Suppose further that you guessed he would reject all of your counteroffers below $G
How much do you earn for such a guess? 0.5
Continue
Control Question
Suppose your proposition was to keep $XXX for yourself and to offer $10-XXX to Player B. Suppose further that you g
that Player B will reject this proposition and that he will make a counteroffer, in which he offers $YYY to you. Assume P
indeed rejected such a proposition and made a counteroffer, in which he wants to keep $10-ZZZ for himself.
How much do you earn for your guess? 0.5
Continue

Player A
I accept this offer
I reject and counter offer $__ to Player A
?
A
B
Please guess whether Player B accepts or rejects your offer. In the following, you are asked to make this guess fc offer you could have made. In case you think Player B rejects, we will also ask you to guess how much Player counteroffer you.
Next
Player B
I accept this offer
I reject and counter offer $__ to Player A
?
A
B
Please guess whether you think Player A expects you to accept or to reject his offer. In the following, you are as make this guess for each of his possible offers. In case you think Player A expects you to reject his offer, we will also a to guess how much he expects you to counteroffer him.
Next

Player A
I accept this offer
I reject and counter offer $__ to Player A
?
Suppose you chose to keep 5 for yourself and to offer 5 to Player B.
Do you think Player B accepts or rejects this proposition?
Accept
Reject
Player B
I accept this offer
I reject and counter offer $__ to Player A
?
Suppose Player A's choice was to keep 8 for himself and to offer you 2.
Do you think Player A expects you to accept or reject this offer?
Accept
Reject

Player B
Player A
I accept this offer
I reject and counter offer $__ to Player A
?
You have just said you believe Player B rejects a proposition in which you keep 8 for yourself and give 2 to Player B.
I think in Player B's counteroffer, I will get 5 of the $10.
Next
You have just said you believe Player A expects you to reject a proposition in which he keeps 1 for himself and giv
you. How much do you think he expects you to counteroffer him?
I think Player A expects a counteroffer, in which he gets $ 5 from the $10.
Next


Player A
I accept this offer
I reject and counter offer $__ to Player A
Please guess whether Player B accepts or rejects your offer. In the following, you are asked to make this guess
offer you could have made. In case you think Player B rejects, we will also ask you to guess how much Play
counteroffer you.
Next
BACK
SKIP
RELOAD
Screens\Belief A\beforebeliefA.vm
start new
Player B
I will reject any counter offer below $__
Suppose still Player A's choice was to keep 0 for himself and to offer you 10.
Suppose you rejected. Please guess which is the smallest counteroffer you could have made to Player A so that just accept it.
I think Player A would reject all of my counteroffers in which he gets less than $
Next
BACK
SKIP

Your Payoff
In the first part of the experiment you earned 7.
For your guesses in the second part of the experiment you earned 9.
Your total payoff therefore is 16 (19 with the show-up fee).
Your Payoff
In the first part of the experiment you earned 3.
For your guesses in the second part of the experiment you earned 9.
Your total payoff therefore is 12 (15 with the show-up fee).

Appendix B

Appendix to Chapter 5

B.1 Instructions

General Instructions

General Remarks
Thank you for participating in this experiment on decision-making. Research foundations have provided funds for conducting this research. During the experiment you and the other participants are asked to make a series of decisions. The money you will earn will depend partly on your own choices and the choices of other participants and partly on chance. All payments will be made confidentially and in cash at the end of the experiment. Please consider all expressions as gender neutral.

Please do not communicate with other participants. If you have any questions after we finish reading the instructions please raise your hand and an experimenter will approach you and answer your question in private.

Two Roles
There are two roles in this experiment: **Player 1** and **Player 2**. At the start of the experiment you will be assigned to one of these two roles through a randomized procedure. Your role will then remain the same throughout the experiment. Your role will only be known to you. Each Player 1 will be randomly paired with a Player 2. No one will ever be informed about the identity of the participant you were paired with nor will anybody else be informed about the choices you made.

Earnings
You will receive $5 for arriving in time. Depending on your decisions, the decisions of other participants and chance you will receive an addional amount according to the rules explained below.

Privacy
This experiment is designed such that nobody, including the experimenters and the other participants, will ever be informed about the choices you or anyone else will make in the experiment. Neither your name nor your student ID will appear on any decision form. The only identifying label on the decision forms will be a number that is known only to you. At the end of the experiment, you are asked one-by-one to collect your earnings in an envelope from a person who has no involvement in and no information about the experiment.

The Decision Situation

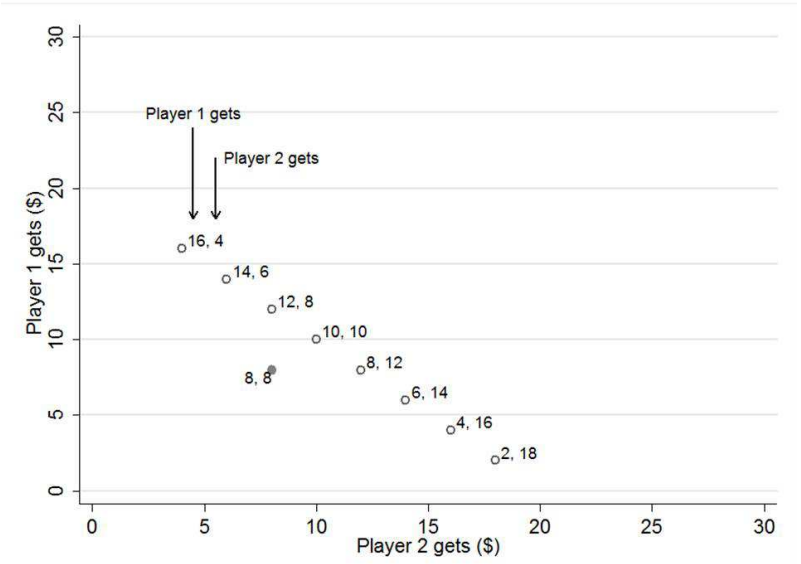
The experiment consists of **60 decision situations**, which are given by graphs. You are asked to choose your preferred option in each of the 60 graphs. Only one graph will be randomly selected for cash payments; thus you should decide which option you prefer in each graph **independently** of your choice in other graphs.

The figure below gives an example of a decision situation. In each situation there are two roles: Player 1 and Player 2.

The first move is made by **Player 1**. He is asked to choose between two options: Option A and Option B.

In each graph **Option A** is a fixed allocation implying a payment for Player 1 and a payment for Player 2. Option A is always represented by a filled dot in the graph. In the graph below Option A implies a payment of \$8 for Player 1 and a payment of \$8 for Player 2.

In each graph **Option B** means that Player 1 gives Player 2 the opportunity to make a choice among a set of possible allocations. Each allocation gives a fixed payment to Player 1 and a fixed payment to Player 2. Option B is always represented by several hollow dots on a line. In the graph below Option B gives **Player 2** the choice between 7 different allocations. For instance, the uppermost point on the line represents an allocation that gives a payment of \$16 for Player 1 and a payment of \$4 for Player 2.



Decision Task - Player 1

If you are assigned the role of Player 1, you are asked to make a choice in each of the 60 graphs between **Option A** (the filled dot assigning you and Player 2 a fixed amount of money) and **Option B** (in which you let Player 2 make a choice between several hollow dots each assigning you and Player 2 an amount of dollars).

Please check one of the boxes below the figure indicating whether you prefer Option A, the filled dot, or Option B, the line of hollow dots.

Decision Task - Player 2

If you are assigned the role of Player 2, you do not know what decision Player 1 is about to make. You are therefore asked - in each of the 60 graphs - to **make a choice as if Player 1 has chosen Option B**, giving you the opportunity to decide on a payoff allocation on the line of hollow dots. The allocation that Player 1 could have chosen is indicated by the filled dot.

Please indicate your choice by circling the preferred allocation on the line of hollow dots.

Earnings

At the end of the experiment one of the 60 decision tasks is chosen randomly and cash payments (in addition to the show up fee of $5) are determined for each pair of participants.

If Player 1 has chosen Option A in that decision task, then Player 1 and the Player 2 paired with this Player 1 will receive the associated payments.

If Player 1 has chosen Option B in that decision task, then the payments for both players depend on the choice made by the paired Player 2. Each of the available choices of the paired Player 2 again implies a payment for both players.

Example: Suppose the graph shown on the previous page is chosen for cash payments in addition to the participation fee. If Player 1 has chosen Option A in this situation than Player 1 receives a payment of $8 and Player 2 a payment of $8. If Player 1 has chosen Option B instead, then the payments of both players depend on the choices of Player 2. Suppose Player 1 has chosen Option B and Player 2 has chosen the uppermost point on the line. Then player 1 receives a payment of $16 and Player 2 a payment of $4.

Control Questions

Question 1: Task of Player 1
Please indicate by a cross which one of the answers about the decision task of Player 1 is true.
☐ Player 1 can choose any of the points on the line with the hollow dots.
☐ Player 1 has no decision to make if Player 2 chooses the filled dot.
☐ Player 1 can choose the filled dot or he can let Player 2 pick one of the hollow dots.
☐ Player 1 can choose any point in the figure.

Question 2: Task of Player 2
Please indicate by a cross which one of the answers about the decision task of Player 2 is true.
☐ Player 2 can choose any of the points on the line with the hollow dots.
☐ Player 2 has no decision to make if Player 1 chooses the filled dot.
☐ Player 2 can choose the filled dot or he can let Player 1 pick one of the hollow dots.
☐ Player 2 can choose any point in the figure.

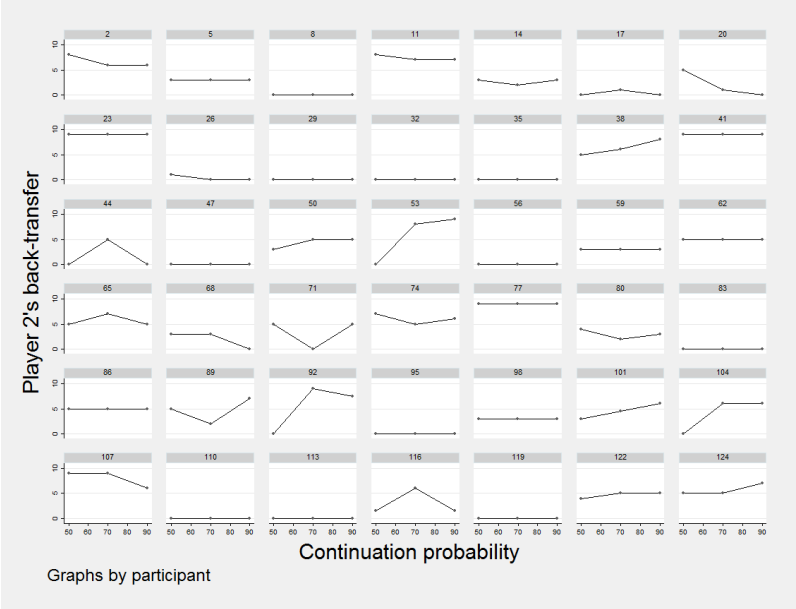
Question 3:
Does Player 2 observe the decision Player 1 has made?
☐ Yes
☐ No

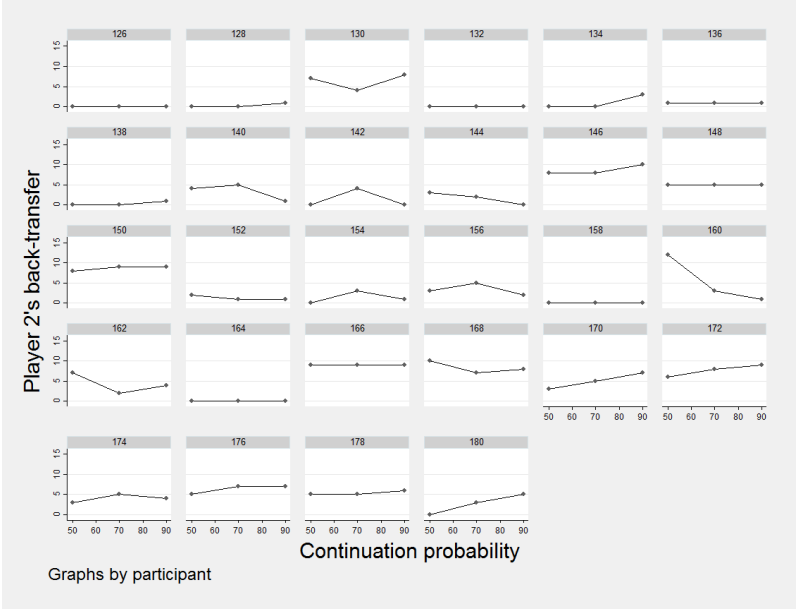
Question 4: Earnings
Suppose Player 2 has chosen the lowermost instead of the uppermost point in the example graph above. Further suppose that the Player 1 paired with this Player 2 has chosen Option A, the filled dot. If this particular decision task was chosen for cash payments, how much would the two players earn (in addition to the show up fee)?
Player 1 would earn $ _____
Player 2 would earn $ _____

Appendix C

Appendix to Chapter 6

C.1 SMs' Individual Back-transfers





C.2 Instructions

General Instructions

General Remarks
Thank you for participating in this experiment on decision-making. During the experiment you and the other participants are asked to make a series of decisions.
Please do not communicate with other participants. If you have any questions after we finish reading the instructions please raise your hand and an experimenter will approach you and answer your question in private. Please consider all expressions as gender neutral.

Three Roles
There are three roles in this experiment: **Player 1**, **Player 2** and the **Observer**. At the start of the experiment you will be assigned to one of these three roles through a random procedure. Your role will then remain the same throughout the experiment. Your role will only be known to you.

Earnings
Depending on your decisions, the outcomes of some random moves and the decisions of other participants you will receive money according to the rules explained below. All payments will be made confidentially and in cash at the end of the experiment.

Privacy
This experiment is designed such that nobody, including the experimenters and the other participants, will ever be informed about the choices you or anyone else will make in the experiment. Neither your name nor your student ID will appear on any decision form. The only identifying label on the decision forms will be a number that is only known to you. At the end of the experiment, you are asked to collect your earnings in an envelope one-by-one from a person who has no involvement in and no information about the experiment.

Decisions Per Period

The experiment is divided into **three periods**. You are asked to choose your preferred option in each of these periods. Only one period will be randomly selected for cash payments; thus you should decide which option you prefer in the given period **independently** of the choices you make in the other periods.

There are three roles in the experiment: Player 1, Player 2 and an Observer.

Player 1 and Player 2

In each period, Player 1 is randomly matched with one Player 2 but none of the participants will interact with the same other participant twice and no one will ever be informed about the identity of the participant he was paired with. Both players receive an endowment of \$10 in each period.

The first move is made by **Player 1**. He is asked to choose whether he wants to send \$3 of his endowment to Player 2 or not.

If Player 1 decides to transfer \$3 to Player 2, his transfer will be multiplied by 5 while being sent. After Player 2 has received the \$15, it is randomly determined whether the round is stopped at this point of time or if Player 2 has the opportunity to send money back to Player 1:

- With the probability $1 - p$, the round continues.
  In this case, **Player 2** can decide how much money he wants to send back to Player 1. He can choose any amount between \$0 and \$15. Player 1 then receives his remaining \$7 plus Player 2's back-transfer as a payment. Player 2 earns his initial endowment (\$10) plus the multiplied transfer (\$15) minus the amount he has chosen to send back to Player 1.
- With a probability p, the round is stopped.
  In this case, Player 1 receives the \$7 that are left from his initial endowment and Player 2 receives his initial endowment (\$10) plus the by five multiplied transfer of Player 1 (\$15).

If Player 1 decides not to transfer the \$3 to Player 2, nothing happens and both players receive their initial endowment of \$10.

The stopping probability p can take values of 10%, 30% or 50%. The realization of p will be stated to all players at the beginning of each period.

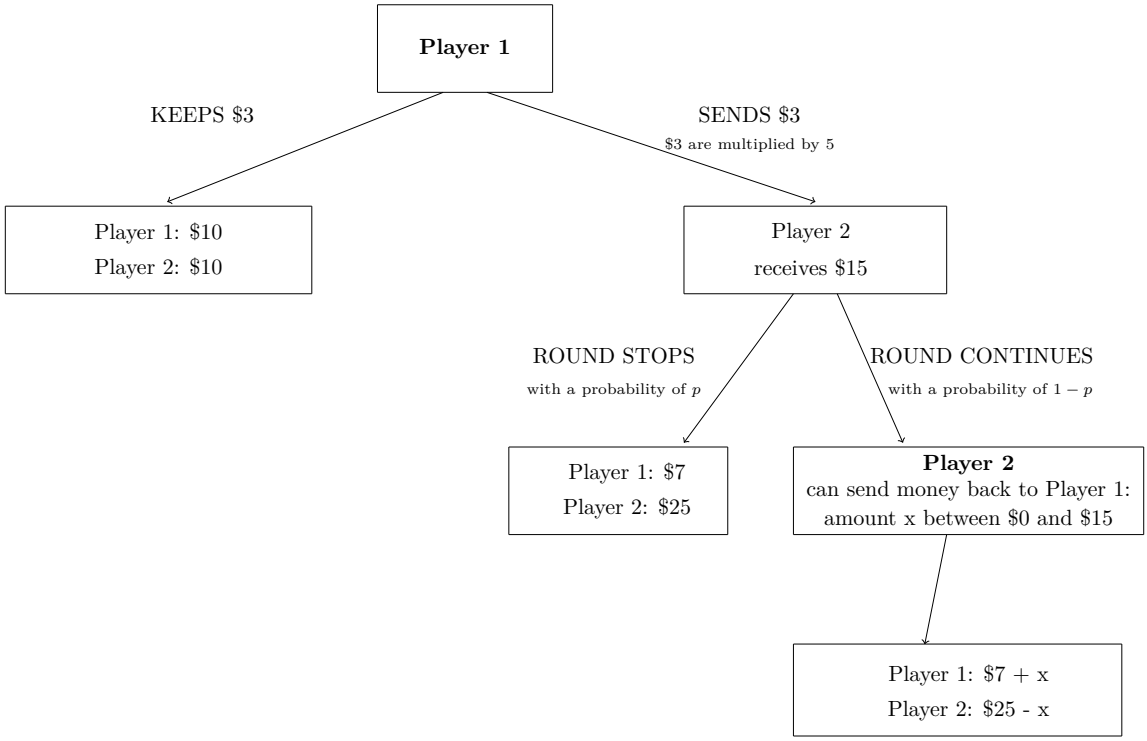
The decision procedure for Player 1 and Player 2 is illustrated by the graph on the following page.

Decision Task Player 1
If you are assigned the role of Player 1, you are asked to choose – in each of the three periods – whether or not to transfer \$3 to Player 2.

Decision Task Player 2
If you are assigned the role of Player 2, you do not know what decision Player 1 is about to make nor what the outcome of the random draw will be. You are therefore asked to decide on how much money you would like to back-transfer to Player 2 assuming Player 1 transferred the \$3 to you and the game was not stopped by the random draw. In each of the three periods, you can choose any amount between \$0 and \$15.

Information Disclosure
At the end of the experiment, one of the periods will be chosen randomly to calculate the cash payments. For this particular period, both players learn whether Player 1 made the transfer of \$3. If he did, it is determined whether the round stops according to the stopping probability p of the chosen period. If the round is not stopped, both players also learn Player 2's decision about his back-transfer.

Decision Stages Player 1 and Player 2



The Observer

In each period, the Observer is asked to guess how much money the participants in the role of Player 2 send on average back to Player 1 assuming that Player 1 transferred the $3 and the random draw allows Player 2 to send money back (the round is not stopped).

Earnings

At the end of the experiment, only one of the periods will be chosen randomly to calculate the cash payments. The exact payments are determined according to the choices that were made and the stopping probability.

Earnings – Player 1 and Player 2
The table below summarizes the payoffs for Player 1 and Player 2 depending on their respective choices.

Choice Player 1	**Random Draw**	**Choice Player 2**	**Payoff Player 1**	**Payoff Player 2**
no transfer	-	-	\$10	\$10
transfer	game continues	back-transfer \$x	\$7 + \$x	\$25 - \$x
	game stops	-	\$7	\$25

Earnings – Observer
The Observer earns money depending on the accuracy of his guess. His payment depends on how much his guess differs from the (rounded) average of all Player 2s' actual choices on the back-transfer in the randomly selected period. The payoffs are summarized in the table below.

Deviation from the average stated back-transfers	**Observer's Payoff**
\$0	\$15
\$1	\$14.5
\$2	\$13
\$3	\$10.5
\$4	\$7
\$5	\$2.5
>\$5	\$0

C.3 Screenshots – Experiment

Introduction

Thank you for participating in this experiment. The purpose of this experiment is to study how people make decisions. In case you should have questions at any time, please raise your hand. Please do not speak to other participants during the experiment and please turn off your mobile phone now. We also ask you not to reveal any details about the experiment after you have participated.

Your payment depends on the decisions you make in the experiment. It is therefore important that you pay attention to the instructions and make your choices thoughtfully. At the end of the experiment, you can collect your payment in cash privately in a sealed envelope from the Economics and Finance front office.

Please enter your 5-digit participant number here: 11111

Continue

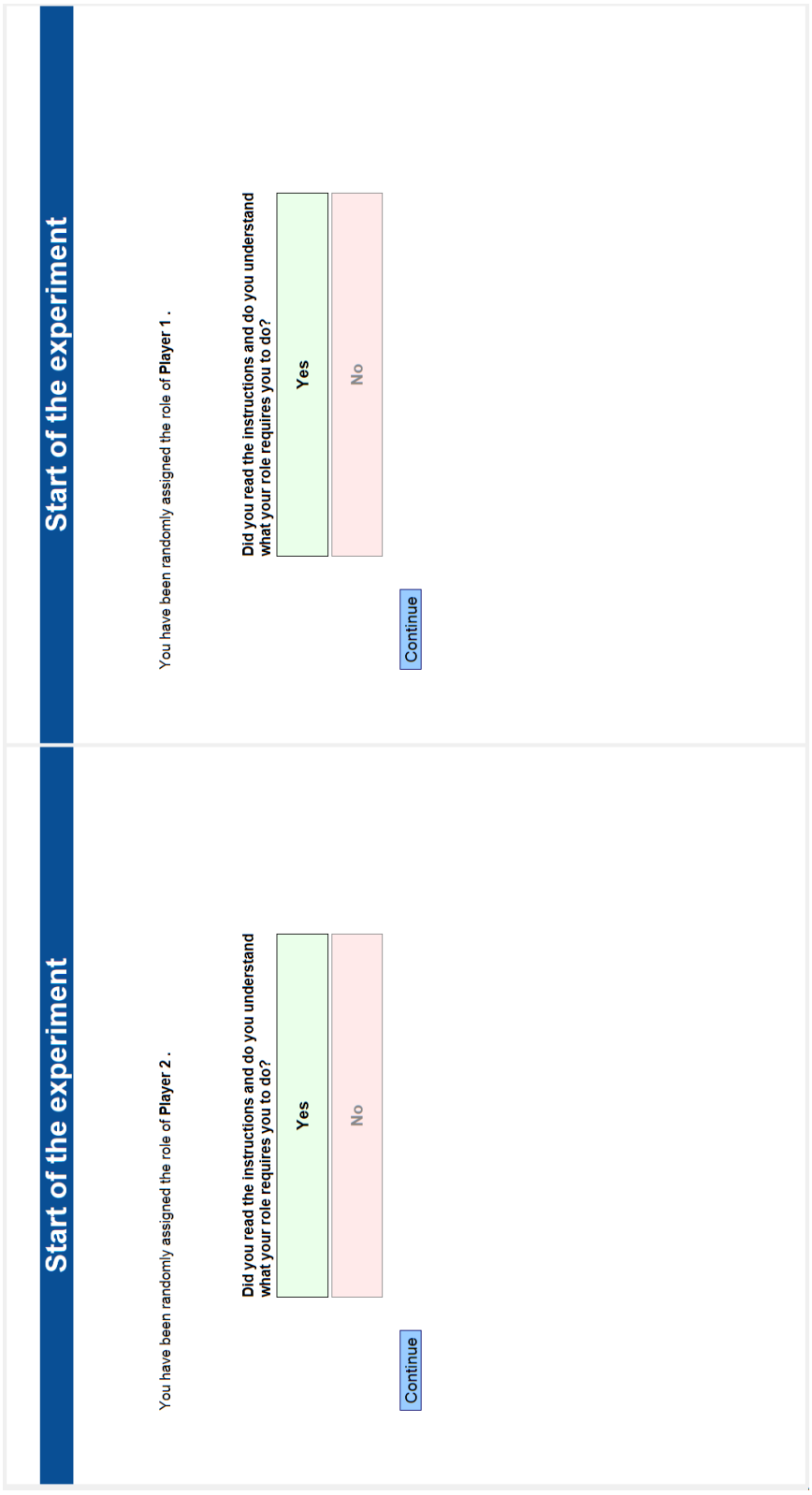
Introduction

Thank you for participating in this experiment. The purpose of this experiment is to study how people make decisions. In case you should have questions at any time, please raise your hand. Please do not speak to other participants during the experiment and please turn off your mobile phone now. We also ask you not to reveal any details about the experiment after you have participated.

Your payment depends on the decisions you make in the experiment. It is therefore important that you pay attention to the instructions and make your choices thoughtfully. At the end of the experiment, you can collect your payment in cash privately in a sealed envelope from the Economics and Finance front office.

Please enter your 5-digit participant number here: 22222

Continue


Start of the experiment
You have been randomly assigned the role of Player 1 .
Did you read the instructions and do you understand what your role requires you to do?
Yes
No
Continue
Start of the experiment
You have been randomly assigned the role of Player 2 .
Did you read the instructions and do you understand what your role requires you to do?
Yes
No
Continue

Practice Questions
1. Suppose the stopping probability was 30%. So in 70 out of 100 times, the game continues if Player 1 decides to transfer the $3 to Player 2. Suppose further that all Player 1s indeed chose to make this transfer and the average back-transfer by Player 2 was $X. Which of the following statements is correct?
In addition to their remaining $7, all Player 1s earn on average 0.7*$X.
In addition to their remaining $7, all Player 1s earn on average $X.
In addition to their remaining $7, all Player 1s earn on average $0.
2. Suppose now that the stopping probability was 50%. So in around 50 out of 100 times, the game continues if Player 1 decides to transfer the $3 to Player 2. Suppose further that all Player 1s again chose to make this transfer and the average back-transfer by Player 2 was $X. Which of the following statements are correct?
All Player 1s earn on average less than if the stopping probability was 30%.
All Player 1s earn on average more than if the stopping probability was 30%.
All Player 1s earn on average more than if the stopping probability was 10%.
Next
Practice Questions
1. Suppose the stopping probability was 30%. So in 70 out of 100 times, the game continues if Player 1 decides to transfer the $3 to Player 2. Suppose further that all Player 1s indeed chose to make this transfer and the average back-transfer by Player 2 was $X. Which of the following statements is correct?
In addition to their remaining $7, all Player 1s earn on average 0.7*$X.
In addition to their remaining $7, all Player 1s earn on average $X.
In addition to their remaining $7, all Player 1s earn on average $0.
2. Suppose now that the stopping probability was 50%. So in around 50 out of 100 times, the game continues if Player 1 decides to transfer the $3 to Player 2. Suppose further that all Player 1s again chose to make this transfer and the average back-transfer by Player 2 was $X. Which of the following statements are correct?
All Player 1s earn on average less than if the stopping probability was 30%.
All Player 1s earn on average more than if the stopping probability was 30%.
All Player 1s earn on average more than if the stopping probability was 10%.
Next

Decision Player 2

In this period, the probability that the game stops after Player 1 made the transfer is 30%. This means that Player 1 receives your back-transfer in 70% of the time and in 30% he earns his remaining $7.

Assume Player 1 transferred you the $3 and the game has not stopped so that you can send money back to Player 1.

How many of the received $15 would you want to send back to Player 1? Please enter a number between 0 and 15:

$ 5

Continue

Decision Player 1

In this period, the probability that the game stops after you made the transfer (and Player 2 cannot return any money) is 30%.

What do you want to do?

send $3

keep $3

Continue

Decision Player 2
In this period, the probability that the game stops after Player 1 made the transfer is 50%. This means that Player 1 You can now decide if you want to send $3 to Player 2.
receives your back-transfer in 50% of the time and in 50% he earns his remaining $7.
If you do so your transfer gets multiplied by 5 before reaching Player 2.
Assume Player 1 transferred you the $3 and the game has not stopped so that you can send money back to Player
1.
In this period, the probability that the game stops after you made the transfer (and Player 2 cannot return any money) is 50%.
How many of the received $15 would you want to send back to Player 1? Please enter a number between 0 and
15:
$ 5
Continue
Decision Player 1
What do you want to do?
send $3
keep $3
Continue

Decision Player 2
In this period, the probability that the game stops after Player 1 made the transfer is 10%. This means that Player 1 receives your back-transfer in 90% of the time and in 10% he earns his remaining $7.
Assume Player 1 transferred you the $3 and the game has not stopped so that you can send money back to Player 1.
How many of the received $15 would you want to send back to Player 1? Please enter a number between 0 and 15:
$ 9
Continue
Decision Player 1
In this period, the probability that the game stops after you made the transfer (and Player 2 cannot return any money) is 10%.
What do you want to do?
send $3
keep $3
Continue

Final Payoff

The random draw determined period **1** for payments. The associated probability that the game stopped was **30%**.
In this period, the following choices and random draws were made:

Decision Player 1	Round Stopped	Your Back-Transfer
send	NO	2

Your resulting final payoff is $23.

Continue

Final Payoff

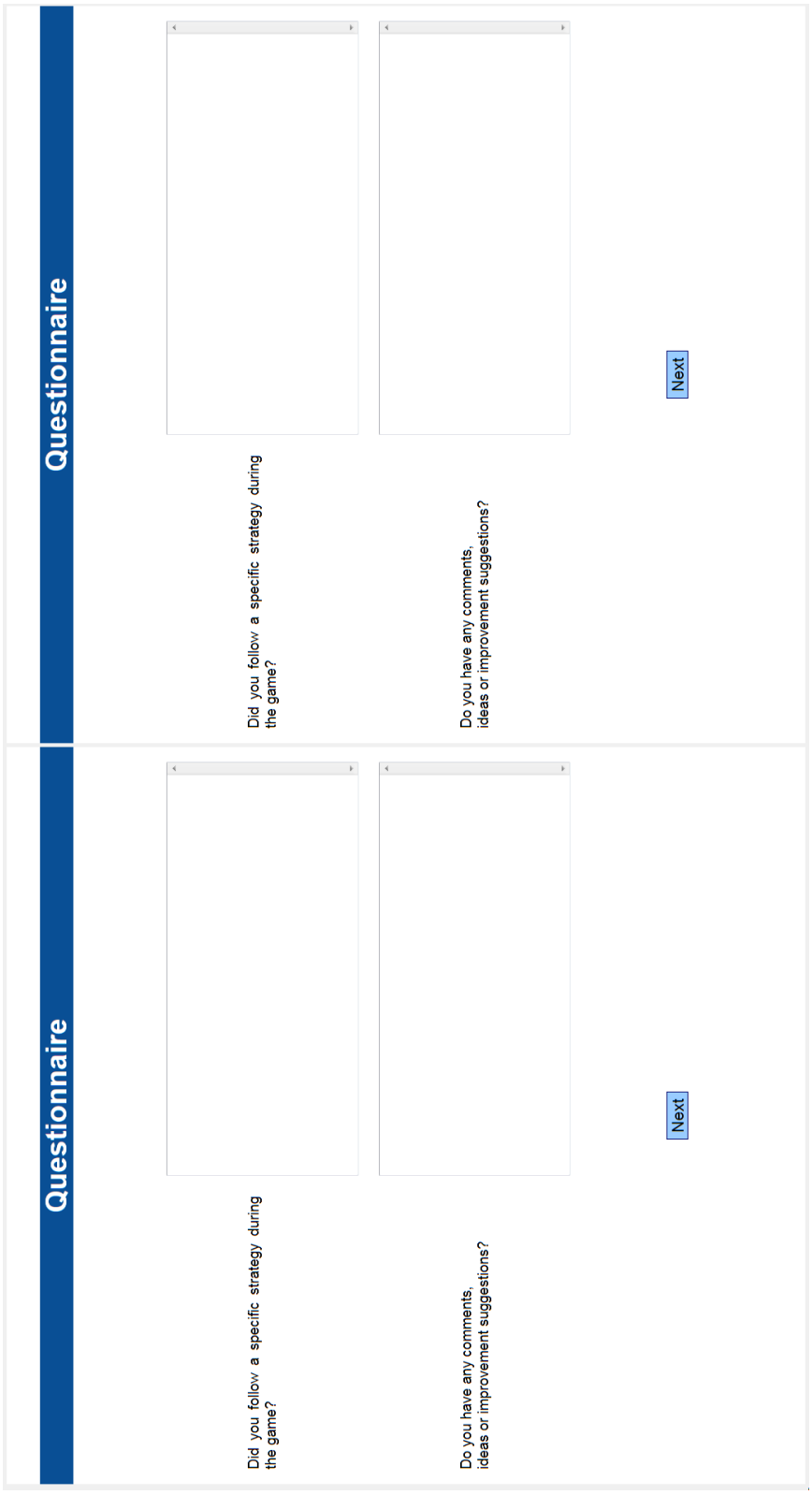
The random draw determined period **2** for payments. The associated probability that the game stopped was **50%**.
In this period, the following choices and random draws were made:

Your Choice	Round Stopped	Back-Transfer Player 2
send	NO	2

Your resulting final payoff is $9.

Continue

Questionnaire
What is your year of birth?
Please indicate your gender.
Male
Female
What is your course of study?
Have you ever participated in economic experiments before?
Yes
No
How many of the participants in this session do you know ?
Where are you from?
Australia
China
India
Europe
USA
Other Asian Country
Other Country
Next
Questionnaire
What is your year of birth?
Please indicate your gender.
Male
Female
What is your course of study?
Have you ever participated in economic experiments before?
Yes
No
How many of the participants in this session do you know ?
Where are you from?
Australia
China
India
Europe
USA
Other Asian Country
Other Country
Next


Questionnaire
Did you follow a specific strategy during the game?
Do you have any comments, ideas or improvement suggestions?
Next
Questionnaire
Did you follow a specific strategy during the game?
Do you have any comments, ideas or improvement suggestions?
Next

FINISHED
The experiment is finished now, please wait for instructions to collect your payment.
FINISHED
The experiment is finished now, please wait for instructions to collect your payment.

Guess Player 3
For this decision round the probability that the game stops after Player 1 made the transfer is 10%.
Assume Player 1 transferred Player 2 the $3 and the game has not stopped. Differently stated: Player 2 can send an amount between $0 and $15 back to Player 1.
How much money do you think the participants in the role of Player 2 send on average back to Player 1?
$
Continue

Guess Player 3
For this decision round the probability that the game stops after Player 1 made the transfer is 30%.
Assume Player 1 transferred Player 2 the $3 and the game has not stopped. Differently stated: Player 2 can send an amount between $0 and $15 back to Player 1.
How much money do you think the participants in the role of Player 2 send on average back to Player 1?
$
Continue

Guess Player 3
For this decision round the probability that the game stops after Player 1 made the transfer is 50%.
Assume Player 1 transferred Player 2 the $3 and the game has not stopped. Differently stated: Player 2 can send an amount between $0 and $15 back to Player 1.
How much money do you think the participants in the role of Player 2 send on average back to Player 1?
$
Continue